

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Magnetic cage  
and flux rope  
constrain the  
nature of solar  
eruptions PAGE 211

## SUN TRAP

SOCIETY

### SCIENCE AFTER APARTHEID

How black South Africans  
are reshaping research

PAGE 159

BIOMECHANICS

### NOT-SO-FAST FOOD

Manoeuvrability trumps  
speed for predators and prey

PAGES 176 & 183

MOLECULAR BIOLOGY

### MISTAKES IN COPIED DNA

Watson and Crick's mutation  
mechanism revisited

PAGES 180 & 195

NATURE.COM/NATURE

8 February 2018 £10

Vol. 554, No. 7691

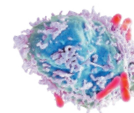


# THIS WEEK

## EDITORIALS

**NUMBER CRUNCHING** Maths gears up for decadal audit of subdisciplines **p.146**

**WORLD VIEW** Broaden the conversation on the use of metrics **p.147**



**CANCER** Mouse study identifies microbes with role in colon tumours **p.149**

## Restore justice in Turkey

*Hundreds of academics and scientists are among those caught up in political crackdowns in Turkey. The government should end the state of emergency.*

Peace is a dangerous cause to fight for in Turkey right now. In the latest blow to academics, 11 members of the Turkish Medical Association, including its president, Raşit Tükel, were arrested in early-morning raids last week. Their crime? Using the slogan that war is a matter of public health, the association had called for a halt to the Turkish army's cross-border assault on military units of Syrian Kurds, launched on 20 January to international consternation. (The Kurdish units targeted have been fighting alongside US troops against the Islamist terrorist organization ISIS in northwest Syria.) The raids follow the arrest of more than a thousand academics who signed a petition in January 2016 calling for peace in the country's southeast, where government forces were fighting Kurdish separatists. Many face criminal charges, and hundreds lost their jobs.

University professors and scientists were also among the 150,000 public servants who were detained and dismissed when draconian laws and a state of emergency were imposed after a failed coup in July 2016. Now, a report published on 18 January (see [go.nature.com/2el9qze](http://go.nature.com/2el9qze)) by human-rights organizations in Turkey shows that many of those dismissed stand accused of supporting the FETÖ, or Gülen, organization believed by the government to have been behind the coup attempt. Membership in other terrorist organizations is also alleged and, as a result, many of these academics face serious terrorism-related charges.

The report details, as far as it is possible to do so, the arrests, detentions and trials of those caught up in the post-coup purges, and raises concern that miscarriages of justice might be occurring on a large scale. Universities have been hard hit — the report says that 5,822 professors and researchers lost their jobs, 380 of whom were signatories of the 2016 Academics for Peace petition. More than 21,000 health-care professionals were among those fired from public service, and a further 4,113 were judges and prosecutors; their loss partly explains why trials are moving forward so slowly. The report notes that even having downloaded a particular encrypted smartphone text-messaging system (called ByLock), favoured by Gülenists and available only through personal introduction, was enough to condemn someone.

The plight of Turkey's academics must not be forgotten. They must be allowed fair hearings and trials without further delay. Telling their stories can be powerful, too. Last week, *Nature* published an interview with theoretical physicist Ali Kaya at Boğaziçi University in Istanbul, who has been charged with being a member of a terrorist organization, about how he managed to carry on his research during his 15 months of incarceration. Colleagues in other countries had tweeted about his achievement — a tactic that other scientists might adopt to help their colleagues in Turkey avoid falling out of the public consciousness.

The general situation in Turkey — whose president is becoming increasingly authoritarian and bellicose, and which hovers on outright civil war — endangers the serious efforts the country has recently been making to improve its research base. As one part of the government

oversees mass arrests and orchestrates war, other parts are quietly but determinedly working to fix some of the entrenched problems in the research system. Thousands of new PhD places have been created in recent years, along with some brand-new research institutes, and uni-

***"The situation in Turkey endangers efforts to improve its research base."***

versities have been energized into competing with each other by offering financial rewards for strong performers. It is a start, and has been enough to persuade at least some young scientists doing postdocs abroad to return home to establish independent research labs.

This is a source of hope in more ways than one. Science can provide a channel for maintaining contact and discussion between countries and cultures in politically tense times. Inevitably, however, like other professionals, many of the scientists making successful careers in Turkey have half-formed emigration plans in mind. The Turkish government needs to make its scientists feel safe. It should revoke the newly extended state of emergency, which has long since outlived its legitimate purpose. ■

## Hardware upgrade

*Artificial intelligence is driving the next wave of semiconductor innovations.*

Advances in computing tend to focus on software: the flashy apps and programs that can track the health of people and ecosystems, analyse big data and beat human champions at Go. Meanwhile, efforts to introduce sweeping changes to the hardware that underlies all that innovation have gone relatively unnoticed.

Since the start of the year, the Semiconductor Research Corporation (SRC) — a consortium of companies, academia and government agencies that helps to shape the future of semiconductors — has announced six new university centres. Having watched the software giant Google expand into hardware research on artificial intelligence (AI), the main chip manufacturers are moving to reclaim the territory. As they do so, they are eyeing the start of a significant transformation — arguably the first major shift in architectures since the birth of computing.

This would be important to science: research in fields from astronomy and particle physics to neuroscience, genomics and drug discovery would like to use AI to analyse and find trends in huge sets of data. But this places new demands on traditional computer hardware. The conventional von Neumann architecture keeps data-storage units inside computers separate from data-processing units. Shuttling information

back and forth between them takes time and power, and creates a bottleneck in performance.

To take advantage of AI technology, hardware engineers are looking to build computers that go beyond the constraints of von Neumann design. This would be a big step forward. For decades, advances in computing have been driven by scaling down the size of the components, guided by Gordon Moore's prediction that the number of transistors on a chip roughly doubles every two years — which generally meant that processing power did the same.

Modern computers bear little resemblance to early machines that used punch cards to store information and mechanical relays to perform calculations. Integrated circuits now contain transistors so small that more than 100 million of them would fit on the head of a pin. Yet the fundamental design of separate memory and processing remains, and that places a limit on what can be achieved.

One solution could be to merge the memory and processing units, but performing computational tasks within a memory unit is a major technical challenge.

Google's AlphaGo research shows a possible, different, way forward. The company has produced new hardware called a tensor processing unit, with an architecture that enables many more operations to be performed simultaneously. This approach to parallel processing significantly increases the speed and energy efficiency of computationally intensive calculations. And designs that relax the strict need to perform exact and error-free computation — a change in strategy known as approximate computing — could increase these benefits further.

As a result, the power consumption of AI programs such as AlphaGo has improved dramatically. But increasing the energy efficiency of such hardware is essential for AI to become widely accessible.

The human brain is the most energy-efficient processor around, so

it is natural for hardware developers to try to mimic it. An approach called neuromorphic computing aims to do just that, with technologies that seek to simulate communication and processing in a biological nervous system. Several neuromorphic systems have already demonstrated the ability to emulate collections of neurons on tasks such as pattern recognition.

These are baby steps, and now the SRC has stepped in to try to encourage the hardware to walk. Under its Joint University Microelectronics

**“The fundamental design of separate memory and processing places a limit on what can be achieved.”**

Program, the SRC has quietly placed its focus on developing hardware architecture. A new centre at Purdue University in West Lafayette, Indiana, for example will research neuromorphic computing, and one at the University of Virginia in Charlottesville will develop ways of harnessing computer memory for extra processing power.

This technological task is huge. So it is heartening to see the SRC, traditionally US-centric, opening its doors. South Korean firm Samsung joined in late 2017, the fifth foreign company to sign up in the past two years. This is a welcome sign of collaboration. But that commercial rivals would work together in this way also signals how technically difficult the industry thinks it will be to develop new hardware systems.

As this research develops, *Nature* looks forward to covering progress and publishing results. We welcome papers that will enable computing architectures beyond von Neumann, such as components for neuromorphic chips and in-memory processing. Scientists across many fields are waiting for the result: computers powerful enough to sift all of their new-found data. They will have to wait a while yet. But the wait should be worth it. ■

# Maths revision

*A decadal update of academic mathematics shows the value of taking one's time.*

Mathematics has its own way of doing things. Not for mathematicians the breakneck chase after the latest academic fad. “It goes up and down over the centuries,” said one expert, when asked whether fluid dynamics — her focus — is now trendy.

Maths moves at its own pace, and the field is currently involved in a global effort to analyse, audit and agree new classifications of how mathematicians study and make use of maths. The MSC2020 system, due to appear in 2020, will formally approve new categories of maths, and split existing definitions into finer classes.

MSC stands for Mathematics Subject Classification, and it provides taxonomical order. In the current MSC2010, for instance, the code 03 represents mathematical logic and foundations. Going deeper, 03E is set theory and 03E72 is fuzzy set theory.

Why bother? The system is jointly managed by the mathematical resource zbMATH, curated by the Leibniz Institute for Information Infrastructure in Karlsruhe, Germany, and by the American Mathematical Society's *Mathematical Reviews*. Each is a ‘meta-journal’ that systematically summarizes and reviews every paper that comes out in the peer-reviewed mathematical literature. *Mathematical Reviews* and zbMATH use the MSC in their internal workflows, and many other journals have adopted the system to assign submissions to editors and reviewers. Mathematicians also use the numerical codes to search for papers in their speciality.

To keep the system up to date, every ten years the two organizations consult reviewers and request suggestions for new entries from the broader community. Nominations opened in July 2016 and close this

August. A theme emerging for proposed new categories is for fields that mix traditional disciplines — such as ‘algebraic statistics’ and ‘numerical algebraic geometry’.

Take topological data analysis, a popular candidate for inclusion. The theory has its roots in topology — the study of shapes and their arrangements within one another — which includes knot theory and higher-dimensional spaces. For more than a century, topology was mostly a pure-maths affair. But researchers have found ways to use it to give structure to large data sets, and so topological data analysis has been born.

More generally, the revision takes the pulse of broader cultural shifts. Suggested new categories indicate that more mathematicians have started to collaborate with researchers in other fields.

Recognition of a new subfield can depend on building citations, and that is a slow process in maths. A recent study of some 20 million references for more than 900,000 mathematical articles in zbMATH found that the time it takes for a paper's citations to peak is several years longer than in other fields — and is lengthening. Consequently, it takes a while for even the most dramatic breakthroughs to register in the MSC system. Many mathematicians expect Peter Scholze, a number theorist at the University of Bonn in Germany, to win a Fields Medal this year for his pioneering work on perfectoid spaces. But, as a research category, perfectoid spaces — only around since 2010 or so — is probably too undercooked yet to make the cut for MSC2020.

Can such a rigid hierarchy survive in an age of fluid metadata and keyword tagging? For now, it remains relevant. Studies have found a high correlation between clustering of the mathematical literature into topics — as measured from citation networks — and the MSC, at least at its upper levels. But things might change. For its own journals, for example, the American Physical Society changed in 2016 from a system similar to the MSC to a hybrid one called Physics Subject Headings. This has both a hierarchical tree of subfields and a broader set of ‘facets’ that cut across them like a Venn diagram, encompassing many terms. Maths might do the same at some point — but, quite correctly, in its own time. Maths has no need to start following fashion now. ■

### **CORRECTION**

The Editorial 'Maths revision' (*Nature* **554**, 146; 2018) mistranslated the name of the Leibniz Institute. It is actually the Leibniz Institute for Information Infrastructure.

DAVE GUTTRIDGE



## Words were a good start — now it is time for action

*Five years ago, the Declaration on Research Assessment was a rallying point. It must now become a tool for fair evaluation, urges Stephen Curry.*

**D**eclarations are bound to fall short. The 240-year-old United States Declaration of Independence holds it self-evident that “all men [*sic*] are created equal”, but equality remains a far-off dream for many Americans.

The San Francisco Declaration on Research Assessment (DORA; <https://sfedora.org>) is much younger, but similarly idealistic. Conceived by a group of journal editors and publishers at a meeting of the American Society for Cell Biology (ASCB) in December 2012, it proclaims a pressing need to improve how scientific research is evaluated, and asks scientists, funders, institutions and publishers to forswear using journal impact factors (JIFs) to judge individual researchers.

DORA's aim is a world in which the content of a research paper matters more than the impact factor of the journal in which it appears. Thousands of individuals and hundreds of research organizations now agree and have signed up. Momentum is building, particularly in the United Kingdom, where the number of university signatories has trebled in the past two years. This week, all seven UK research councils announced their support.

Impact factors were never meant to be a metric for individual papers, let alone individual people. They're an average of the skewed distribution of citations accumulated by papers in a given journal over two years. Not only do these averages hide huge variations between papers in the same journal, but citations are imperfect measures of quality and influence. High-impact-factor journals may publish a lot of top-notch science, but we should not outsource evaluation of individual researchers and their outputs to seductive journal metrics.

Most agree that yoking career rewards to JIFs is distorting science. Yet the practice seems impossible to root out. In China, for example, many universities pay impact-factor-related bonuses, inspired by unwritten norms of the West. Scientists in parts of Eastern Europe cling to impact factors as a crude bulwark against cronyism. More worryingly, processes for JIF-free assessment have yet to gain credibility even at some institutions that have signed DORA. Stories percolate of research managers demanding high impact factors. Job and grant applicants feel that they can't compete unless they publish in prominent journals. All are fearful of shrugging off the familiar harness.

So, DORA's job now is to accelerate the change it called for. I feel the need for change whenever I meet postdocs. Their curiosity about the world and determination to improve it burns bright. But their desires to pursue the most fascinating and most impactful questions are subverted by our systems of evaluation. As they apply for their first permanent positions, they are already calculating how to manoeuvre within the JIF-dependent managerialism of modern science.

There have been many calls for something better, including the Leiden Manifesto and the UK report ‘The Metric Tide’, both released in

2015. Like DORA, these have changed the tenor of discussions around researcher assessment and paved the way for change.

It is time to shift from making declarations to finding solutions. With the support of the ASCB, Cancer Research UK, the European Molecular Biology Organization, the biomedical funder the Wellcome Trust and the publishers the Company of Biologists, *eLife*, F1000, Hindawi and PLOS, DORA has hired a full-time community manager and revamped its steering committee, which I head. We are committed to getting on with the job.

Our goal is to discover and disseminate examples of good practice, and to boost the profile of assessment reform. We will do that at conferences and in online discussions; we will also establish regional nodes across the world, run by volunteers who will work to identify and address local issues.

IT'S WORTH  
DOING THE  
**EXPERIMENT**  
TO PROPERLY  
EVALUATE  
**EVALUATION.**

This week, for example, DORA is participating in a workshop at which the Forum for Responsible Metrics — an expert group established following the release of ‘The Metric Tide’ — will present results of the first UK-wide survey of research assessment. This will bring broader exposure to what universities are thinking and doing, and put the spotlight on instances of good and bad practice.

We have to get beyond complaining, to find robust, efficient and bias-free assessment methods. Right now, there are few compelling options. I favour concise one- or two-page ‘bio-sketches’, similar to those rolled out in 2016 by the University Medical Centre Utrecht in the Netherlands.

These let researchers summarize their most important research contributions, plus mentoring, societal engagement and other valuable activities. This approach could have flaws. Perhaps it gives too much leeway for ‘spin’. But, as scientists, surely we can agree that it's worth doing the experiment to properly evaluate evaluation.

This is hard stuff: we need frank discussions that grind through details, with researchers themselves, to find out what works and to forestall problems. We need to be mindful of the damage wrought to the careers of women and minorities by bias in peer review and in subjective evaluations. And we need to join in with parallel moves towards open research, data and code sharing, and the proper recognition of scientific reproducibility.

Declarations such as DORA are important; credible alternatives to the status quo are more so. True success will mean every institution, everywhere in the world, bragging about the quality of their research-assessment procedures, rather than the size of their impact factors. ■

**Stephen Curry** is a professor of structural biology and assistant provost for equality, diversity and inclusion at Imperial College London. He is also chair of the DORA steering group.  
e-mail: [s.curry@imperial.ac.uk](mailto:s.curry@imperial.ac.uk)

# SEVEN DAYS

The news in brief

## POLICY

### Ivory trade banned

Hong Kong lawmakers voted on 31 January to ban the trading of ivory in the Chinese territory, which is the world's largest ivory market. The ban will be implemented in phases, and will ultimately require traders to dispose of their stock by 2021. Conservationists hailed the move as a victory for elephant preservation. Although ivory sales are banned in most of the world under a 1990 treaty, sales of antiques made of the material have remained legal in Hong Kong, providing cover for illegal trade in fresh ivory. A ban on the substance in mainland China — the biggest market for Hong Kong ivory — came into effect on 31 December last year.

## PEOPLE

### Investigator death

Esmond Bradley Martin, a veteran investigator in the fight against ivory poaching in Africa, was found dead at his home in Nairobi on 4 February. Media reports said that the 76-year-old had been stabbed in the neck. Kenyan police have arrested four people in connection with the death, although the motive for the crime remains unclear. Bradley Martin, a US citizen, had spent decades investigating the trade in elephant tusks and rhino horn across Africa and Asia. His work influenced China's decision to end its legal rhino-horn trade in 1993, as well as the country's ban on ivory sales, which came into force on 31 December 2017. China's ivory trade has been widely blamed for driving elephant poaching; in recent years, Bradley Martin showed that illicit trade

was moving from China to neighbouring countries, such as Laos and Vietnam.

### Astronomy move

Astrophysicist Christian Ott, who in 2015 was found by the California Institute of Technology in Pasadena to have committed gender-based harassment against two graduate students there, is moving to the University of Turku in Finland. "Dr Ott's past was known during the recruitment process, and the matter has been carefully considered," the University of Turku said in a 1 February statement, noting that Ott will not have supervisory responsibilities. The position is for 2 years, with a 4-month

trial period. His appointment has drawn criticism from other astronomers, who say they are concerned about the message the hiring sends to those who have been harassed. Ott was a visiting professor at Kyoto University from March to June 2017, before leaving Caltech last December. Through his lawyer, Ott declined *Nature's* request to comment on the matter.

### Health chief resigns

The head of the US Centers for Disease Control and Prevention (CDC) resigned on 31 January, soon after coming under fire in the press for trading stock in tobacco companies while leading the agency. Brenda

the ground. A truck driver was detained later that day, but subsequently released by a judge; prosecutors have appealed against the ruling. The Nazca Lines were constructed between 500 BC and AD 500, and are thought to have been used in astronomy. They have been designated a World Heritage Site by the United Nations.



GENRY BAUTISTA/AGENCIA ANDINA/EPA

## Truck tracks harm ancient Peruvian site

Peru's ancient Nazca Lines have been damaged by a truck that drove over the cultural site without permission, the country's ministry of culture said on 29 January. The incident, which occurred on 27 January, left deep tracks over an area roughly 50 metres by 100 metres and affected three geoglyphs — images scratched into

Fitzgerald had served as CDC director since last July. On 30 January, Politico broke a story about her tobacco holdings. A spokesman for the US Department of Health and Human Services said in a statement that Fitzgerald's financial holdings required "a broad recusal" that would limit her ability to perform her job.

## FACILITIES

### Neutrino test axed

Italy's physics agency, the INFN, has cancelled a planned neutrino experiment. The Short distance neutrino Oscillations with Borexino (SOX) experiment, a collaboration with the CEA, France's nuclear agency,

SHATIL & ROZINSKI/NPL was designed to determine whether there is a fourth, 'sterile' type of neutrino in addition to the three known ones. The partnership had planned to install a high-intensity neutrino source at the Gran Sasso underground physics laboratories in central Italy, which is home to the Borexino neutrino detector. But Mayak, a Russian firm contracted to make the source, has said that it will be unable to extract enough cerium-144 from nuclear waste for the 18-month experiment. SOX spokesperson Marco Pallavicini says that the agencies had spent about one-third of the estimated €6-million (US\$7.5-million) cost of the experiment.

## ENVIRONMENT

## Wolf lawsuit

Conservation groups filed a lawsuit on 30 January to compel the US Fish and Wildlife Service to strengthen its plan to rescue the endangered Mexican wolf (*Canis lupus baileyi*) from extinction. The animal (pictured) is a rare subspecies of grey wolf; an estimated 150 individuals roam New Mexico, Arizona and northern Mexico. The lawsuit alleges that the federal agency failed to take steps to help the canines recover.



Leading wolf biologists have recommended expanding the animal's range and establishing new populations in the southwestern United States. The groups also claim that the plan inadequately addresses high levels of inbreeding in the Mexican-wolf population, a significant threat to its future existence.

## Satellite launch

Italy and China have launched a satellite that will monitor electromagnetic phenomena from space that may be linked to earthquakes and other seismic activity on Earth. On 2 February, a rocket carrying Zhangheng-1, also called the China Seismo-Electromagnetic Satellite, lifted off from the Jiuquan Satellite Launch Centre in the Gobi Desert in Inner Mongolia. The satellite is equipped with

nine scientific instruments. Data they collect will be used to develop new methods of studying earthquakes from space, say researchers involved in the project. Zhangheng-1 is expected to be in orbit for five years.

## EVENTS

## Fertility licences

Two women have been approved to be the first in the United Kingdom to receive an *in vitro* fertilization procedure called mitochondrial replacement therapy (MRT), which uses the DNA of three people. MRT reduces the risk of women passing on certain inherited diseases caused by mutations in mitochondrial DNA. The United Kingdom legalized the procedure in 2015 after a parliamentary vote, but people wishing to

undergo treatment must be approved individually by the country's fertility-treatment regulator, the Human Fertilisation and Embryology Authority. The procedures will be carried out at Newcastle Fertility Centre. MRT has already been successfully performed in Mexico and Ukraine.

## FUNDING

## Funding boost

India will increase its investment in science by 10% to 536.2 billion rupees (US\$8.4 billion) for 2018–19, compared with the previous year. The budget, released on 1 February, includes 30.7 billion rupees earmarked for a digital programme that includes artificial intelligence and cyber systems. Despite the new money, the country's spending on science will remain relatively low, at around 0.8% of gross domestic product (GDP). Scientists have called for the government to boost investment in science, technology and research to 3% of GDP.

## AI start-up fund

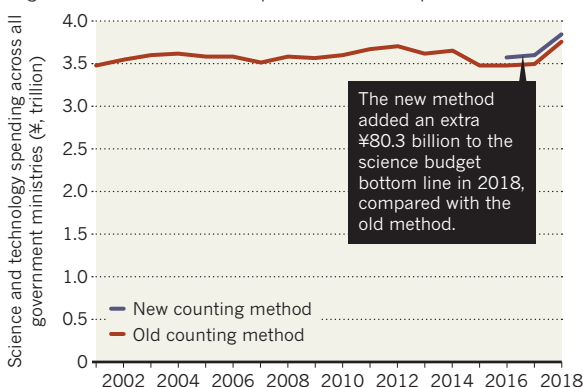
Leading machine-learning researcher Andrew Ng has launched a US\$175-million fund dedicated to nurturing artificial intelligence (AI) start-up companies. Ng, who was previously head of AI at Chinese tech giant Baidu, has raised money from some of Silicon Valley's biggest investors. He wrote in a blog post on 30 January that the effort, called the AI Fund, will aim to build companies from scratch and allow quickly developing AI firms to focus on research rather than on fundraising. In December, Ng launched Landing.AI, a start-up aimed at bringing AI to the manufacturing industry — for example, by developing automated visual-inspection systems to spot defects in products.

## TREND WATCH

Japan will increase its spending on science and technology by 7% to ¥3.84 trillion (US\$35 billion) in 2018 compared with the previous year, the government's science advisory body announced on 30 January. The rise comes after stagnant growth in the science budget since the early 2000s. Prime Minister Shinzo Abe's government aims to boost Japan's science and technology budget by ¥300 billion per year to meet a goal of spending 1% of gross domestic product on research by 2020, up from 0.65% in 2015.

## JAPAN'S SCIENCE SPENDING

This year, the government adopted a standardized method for counting science spending by individual ministries to follow the practices of the Organisation for Economic Co-operation and Development.



# NEWS IN FOCUS

**PUBLISHING** South Korea investigates researchers who put kids as authors **p.154**

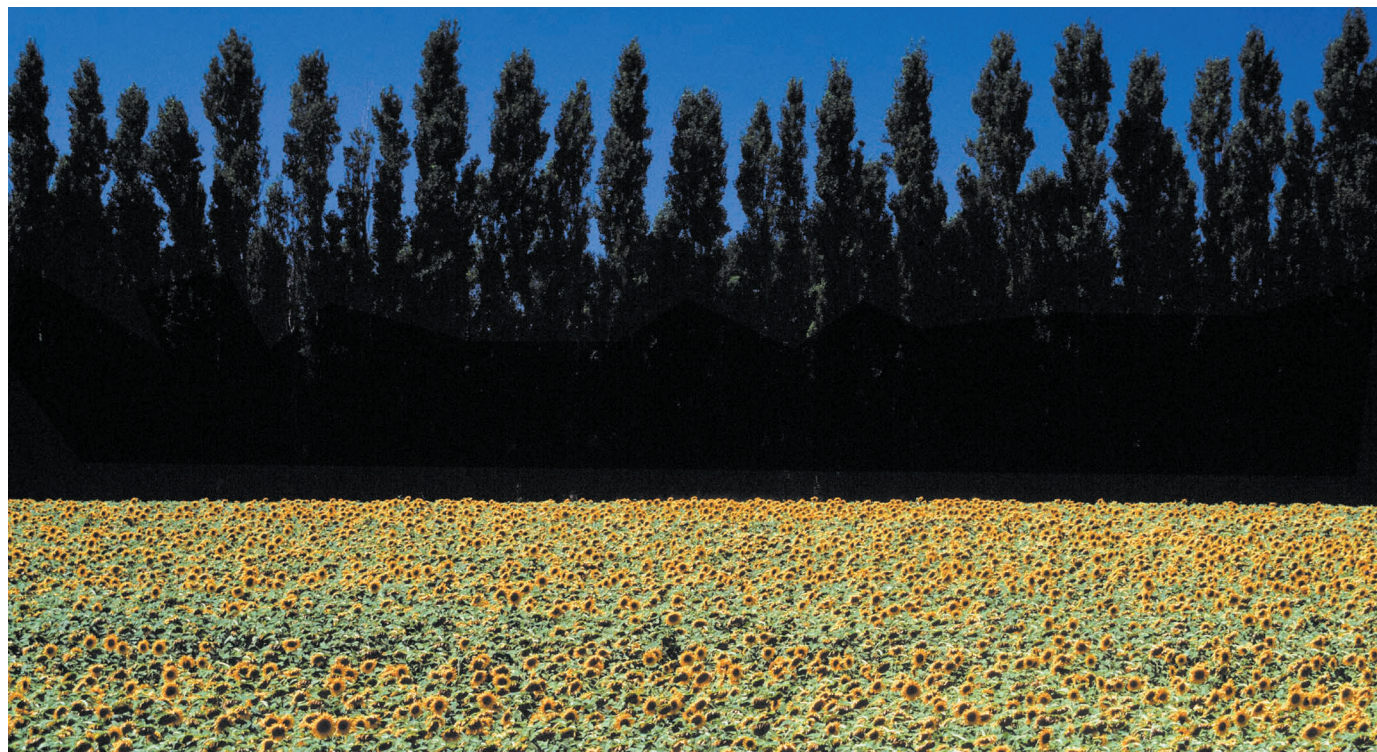
**CLIMATE** Countries consider curbing black-carbon pollution from ships **p.155**

**GENOMICS** Invasive crayfish species evolved through unusual method **p.157**



**SOUTH AFRICA** Researchers grapple with decolonization **p.159**

DENNIS STOCK/MAGNUM



Sunflowers and all other flowering plants probably arose from a common ancestor.

## PLANT SCIENCE

# Debate blooms over Earth's first flower

*Some scientists doubt a statistical prediction of the ancestral blossom's structure.*

BY HEIDI LEDFORD

An ambitious effort to reconstruct the world's first flower has seeded a debate over what that blossom looked like — and, more broadly, which forms a flower can take.

The project, called eFLOWER, combined an unparalleled database of plant traits, reams of molecular data on evolutionary relationships, and complex statistical models to determine what the ancestor of all modern flowering plants might have looked like.

When the study's results were published last August (H. Sauquet *et al. Nature Commun.* **8**, 16047; 2017), they drew intense interest from academics and the media.

But since then, researchers have raised questions about some of eFLOWER's predictions. On 31 January, plant morphologist Dmitry Sokoloff at Moscow State University and his colleagues published a reanalysis of the data that suggests a different arrangement of key female reproductive structures in the first flower (D. Sokoloff *et al. Am. J. Bot.* <http://dx.doi.org/10.1002/ajb2.1003>; 2018).

The debate centres on the finer points of flower architecture, but raises a broader concern about using statistical models and large data sets to tackle biological questions, says Pamela Soltis, a plant biologist at the University of Florida in Gainesville. "Things can be statistically possible without being biologically possible," she says.

Flowering plants are a remarkable evolutionary success. Although they appeared as recently as 140 million years ago — about 200 million years after the first seed plants — they now make up about 90% of all living

► land plants. But fossil flowers are scarce, and botanists have long speculated about what the first blooms might have looked like. “The flower was responsible for this massive diversification,” says Soltis. “We can’t understand how we got to where we are without understanding what the first one was like.”

About eight years ago, the eFLOWER project enlisted a team of botanical experts to find out. The team catalogued more than 20 traits in nearly 800 species. They then matched these data with molecular studies of evolutionary relationships, and used statistical modelling to infer the features of the earliest flower.

### BUDDING DOUBTS

The results painted a picture of a flower that was symmetric around a central axis and contained both male and female sex organs. The eFLOWER models also suggested that many organs in the first flower were whorled, meaning they were arranged regularly in concentric circles when viewed from above. But the authors also warned that statistical support for some of these findings was weak.

Even so, the idea of a whorled ancestral flower shocked some people, says Hervé Sauquet, a lead author on the eFLOWER paper and an evolutionary biologist now at the Royal Botanic Garden in Sydney, Australia. Many plant scientists expected that the bloom’s

organs would have been staggered in a spiral coiled around a central axis. “It was a long-held dogma that was never confirmed,” he says.

But what puzzled Sokoloff was that in Sauquet’s analysis, the flower’s petals and male reproductive parts were arranged in whorls, yet the female reproductive organs, carpels, were arranged in a spiral. He had never seen this combination of whorled and spiral organs in a single flower. Moreover, he and his colleagues suggest that it might not be developmentally possible for plants to achieve two different arrangements of organs in one flower.

That’s because the organs emerge from the same region of the plant, Sokoloff says. In some whorled flowers, the position of the carpels dictates the position of the male reproductive organs. Sokoloff’s team picked back through the eFLOWER database and found four examples in which whorled and spiral organs had been identified within the same flower. But after further analysis, they decided that each example contained only one type of reproductive organ.

Sauquet says that his team has since revisited those data and agreed with some, although not all, of Sokoloff’s concerns. Repeating their

analysis with an updated and expanded data set, they now find that all reproductive organs in the ancestral flower were probably whorled, he says. But some of the revised results had a relatively low degree of statistical support, just as the first analysis did. “It wasn’t certain before, and it remains uncertain,” Sauquet says. “We don’t know the final answer yet.”

Sokoloff says that a fundamental problem of eFLOWER’s approach was evaluating each trait of a flower independently before assembling those traits into a coherent bloom. “They analysed the evolution of each character separately,” he says. “But some combinations of characters are impossible.”

Even so, Sauquet argues that the absence of a particular form in modern flowers does not mean that it never existed. “There are a lot of weird things that existed before that we cannot see nowadays,” he says.

Settling the debate over the first flower will take a bigger database and more-sophisticated models, says Wenheng Zhang, who studies plant evolution at Virginia Commonwealth University in Richmond. But the eFLOWER effort is an example of how modern techniques can be married to classical morphology to tackle fundamental questions about plant origins, she says. “This kind of study redirects botanists to look at the morphology,” Zhang says. “It just comes back to the basics.” ■

### SOUTH KOREA

# Child authors spark probe

*Researchers may have added relatives to papers to boost their chances at university.*

BY MARK ZASTROW

The South Korean government is expanding an investigation into researchers who named their children as co-authors on papers. In some cases, the practice is thought to be intended to give the children an edge when applying to university, a highly competitive process in South Korea. The education ministry announced on 1 February that it would extend its original investigation, which last month identified

82 academic papers on which authors had named their children or relatives — many of them in middle or high school — as co-authors.

And on 4 February, the science ministry launched its own investigation into several of the country’s elite technical universities, which had not been included in the education ministry’s initial probe.

The 82 papers with child authors were uncovered in a month-long review of articles written by more than 70,000 full-time

university staff members across arts and sciences over 10 years. The review was prompted by a single case of child authorship that came to light late last year, at Seoul National University.

The investigation results, released on 25 January, found examples from 29 South Korean universities. In 39 of the papers, the students seemed to have participated in the research as part of a programme related to their school curriculum; the other 43 appeared not to have, according to the investigation.



### IMAGES OF THE MONTH



January’s sharpest science shots, selected by Nature’s photo team [go.nature.com/2nh1pq2](http://go.nature.com/2nh1pq2)

### MORE NEWS

- Physicists harness twisted mathematics to make powerful laser [go.nature.com/2e5qu07](http://go.nature.com/2e5qu07)
- Indian science budget fails to impress [go.nature.com/2gsp3n6](http://go.nature.com/2gsp3n6)
- PubMed Commons closes its doors to comments [go.nature.com/2elkyqa](http://go.nature.com/2elkyqa)

### NATURE PODCAST



Crayfish clones; the social smarts of magpies; and building tougher wood [nature.com/nature/podcast](http://nature.com/nature/podcast)

SEKOURI KAMEL/GETTY IMAGES  
ROBINSON PHOTOGRAPHY  
REV SHUTTERSTOCK

The education ministry has not released the names of the researchers involved in the cases, nor the journals in which they published. However, according to South Korean media reports, many of the papers appeared in journals included in the Science Citation Index (SCI). The ministry told *Nature* that the initial review relied on universities to self-report cases, and was not exhaustive because many staff members were on their winter holidays.

### INDEXED PAPERS

In its continuing probe, the education ministry intends to examine papers by South Korean authors indexed in citation databases, including the SCI, Web of Science and Scopus, and to cross-check the names against the family relationships of 76,000 full-time faculty members. The investigation will run until 16 March.

The ministry intends to refer each case to the corresponding university's research-ethics committee to confirm whether it constitutes misconduct or legitimate authorship. If the student co-authors did not participate in the research, academics will face possible disciplinary action, including dismissal, the ministry said.

So far, the most-affected universities include some of Seoul's elite tertiary institutions: Sungkyunkwan University (eight cases), Yonsei University (seven cases), Seoul National University and Kookmin University (six each). A Sungkyunkwan spokesperson confirmed that the university would be opening probes as per the ministry's request, including possible penalties of dismissal.

Yonsei University declined to answer *Nature's* questions about the investigation, pending further information from the government. A Seoul National University spokesperson emphasized that there has not yet been any finding suggesting that actual misconduct had occurred, and that its research-integrity committee would investigate all cases.

A spokesperson for Kookmin University told *Nature* that an initial review of the institution's cases indicated that the collaborations were legitimate. "We have some records and notes that their children participated in a lot of activities. So we think we don't have any problem," he said.

The practice has sparked a national outcry. In an editorial, the *Korea Herald* called the acts "no less than fraud, which greatly threatens the integrity of universities and education as a whole in Korea".

The education ministry said that any students listed as co-authors who did not participate in the research would have their university admission revoked. ■



Most container ships burn heavy diesel fuel that produces black carbon.

### AIR POLLUTION

# UN targets black carbon from ships

*Nations are advancing efforts to reduce sooty emissions.*

BY JEFF TOLLEFSON

Governments are poised this week to begin discussing rules to curb black-carbon pollution from ships, after nearly seven years of preparation. The sooty emissions, which are produced by diesel engines, warm the climate and harm human health.

At a meeting in London, a panel of the United Nations International Maritime Organization (IMO) is expected to agree on measurement techniques to gather data that could support eventual regulations. That is the second step in a three-step process begun in 2011. Agreeing on a definition for black carbon took four years; the final step, writing rules, could take a few more.

Reducing the amount of black carbon emitted by ships could have a significant impact on the climate. The pollutant, a melange of particles and oil droplets that come in many shapes and sizes, is the second-largest driver of global warming — behind only carbon dioxide. Diesel engines, such as those in ships, account for around one-fifth of the world's black-carbon emissions, according to a study published in

2013 (T. C. Bond *et al.* *J. Geophys. Res. Atmos.* **118**, 5380–5552; 2013).

The pollution is also dangerous when inhaled, in part because black-carbon particles collect other contaminants — such as sulfuric acid and heavy metals — as they travel through the atmosphere. Advocates are pushing the IMO to speed up its negotiations, which involve more than 170 countries. "We really only have 90 minutes per year where we are actively discussing the topic, so it's easy to delay and to stall," says Bryan Comer, a senior researcher at the International Council on Clean Transportation, a non-profit research group in Washington DC.

Although global black-carbon emissions from diesel engines on land are roughly 20 times higher than those from ships' engines, the health and environmental impacts of shipping pollution hits many busy ports and coastal areas disproportionately hard, says Daniel Lack, an independent consultant in Brisbane, Australia. "When you concentrate all of these ships into specific areas, all of a sudden they become one of the most dominant sources of pollution."

One area of special concern is the rapidly melting Arctic. The region's shipping traffic ►

► is projected to increase in the coming decades, as sea ice recedes — a thaw that could be exacerbated by particles of black carbon, which hasten melting when they land on snow and ice.

Measuring black-carbon emissions is not a trivial task, Lack says. The most accurate, and expensive, technology fires a laser pulse through exhaust samples in a tube. Black-carbon particles absorb and then release the energy from the pulse, creating a pressure wave whose strength is equivalent to the amount of light that was absorbed. The shipping industry is pushing for a cheaper, but less accurate, method that draws exhaust through a filter; measurements of the reflectivity of that filter before and after use are then used to determine how much pollution a ship emitted.

Both approaches could serve a purpose as the IMO moves forward with regulations, Comer says. But he adds that many of the regulatory actions that the organization could pursue to reduce black-carbon emissions do not require regular measurements from ships. Shifting from 'heavy' fuel oil to cleaner types — similar to those used in trucks — would reduce ships' black-carbon output by 35–80%, depending on the engine. And installing filters on the vessels' exhaust systems would cut emissions by at least 85%.

The shipping industry is under pressure to curb other types of pollution. The United States, Canada and the European Union already require ships to use lower-sulfur fuels in some coastal zones. And in 2016, the IMO agreed to reduce the sulfur content in all shipping fuels from 3.5% to 0.5% by 2020. That is good news for public health, but it could inadvertently exacerbate global warming, says James Corbett, an engineer at the University of Delaware's School of Marine Science and Policy in Newark.

In a study this week (M. Sofiev *et al.* *Nature Commun.* **9**, 406; 2018), Corbett and his colleagues found that the IMO sulfur standard could reduce global cardiovascular and lung-cancer deaths attributable to fine particulate matter by 2.6%, and the incidence of childhood asthma by 3.6%. But the new standard could accelerate climate change by decreasing the number of bright, sulfur-containing particles in the atmosphere that cool the planet by reflecting sunlight back into space. The researchers estimate that this effect would increase the human contribution to warming by around 3%.

"We're talking some big numbers," says Corbett.

For Comer, that is all the more reason to press forward with black-carbon regulations. "It's frustrating," he says. "We already know how to control [black-carbon] emissions, but we're stuck going through the three-step process." ■



Cuts to Bulgaria's science budget sparked strikes in November, following earlier protests by students.

#### BUDGETS

# Funding cuts hit Bulgarian science

*A push to attract investment in innovation has floundered.*

BY INGA VESPER

European Union science ministers met on 2 February in their bloc's poorest member state — Bulgaria — to discuss future EU research policy. For the host nation, it was supposed to be a chance to showcase ambitious plans to boost economic growth by attracting international research institutes to the country.

But the timing of the event was awkward, to say the least. In July 2017, Bulgaria had been due to receive €150 million (US\$186 million) from the EU to build facilities for research and innovation, under a programme that aims to boost economic growth in poor regions. The programme, which was expected to give Bulgaria €700 million between 2014 and 2020, is designed to help with the costs of research infrastructure.

However, EU authorities withheld the money after Bulgaria failed to identify enough sufficiently qualified scientists to evaluate the proposals. Then, in November 2017, the Bulgarian government cut its 2018 science and higher education budget by around 25%, a move it had planned in anticipation of the windfall.

The decision has frustrated scientists in Bulgaria, because they had wanted to use the new infrastructure to forge links with

researchers outside the country. "Now, we cannot prepare proposals because we are not going to have the infrastructure," says Ana Proykova, a physicist at Sofia University and an adviser on European research infrastructure to Bulgaria's government. She says that the government should reinstate the funds it cut from the 2018 budget. "We are still fighting very strongly for the funding procedure to be reopened, even if it is in the middle of this year. Otherwise, our budget is going to be very tiny."

Bulgaria, which took over the six-month rotating presidency of the EU on 1 January, produces little science compared with the bloc's other member states. The country's output is low (see 'Bulgaria's output lags behind'), and more than 30% of PhD-holding Bulgarians are at present pursuing careers abroad. But scientists in Bulgaria hope for improvements. The country intends to bid for a proposed Balkan synchrotron particle accelerator, a light source that many hope will promote international diplomacy in the region.

Its universities still want to tap into EU infrastructure funds. During its presidency, Bulgaria is also in charge of negotiating Framework 9, the EU's latest seven-year plan for science, which is due to be finalized in May. It sees the plan, in part, as an opportunity for Bulgarian companies to enter into

OLEG POPOV/REUTERS

lucrative contracts with international research consortia. “Industry is very important for us,” says Karina Angelieva, adviser for education and research at Bulgaria’s permanent representation to the EU, in Brussels.

### RAISED SCRUTINY

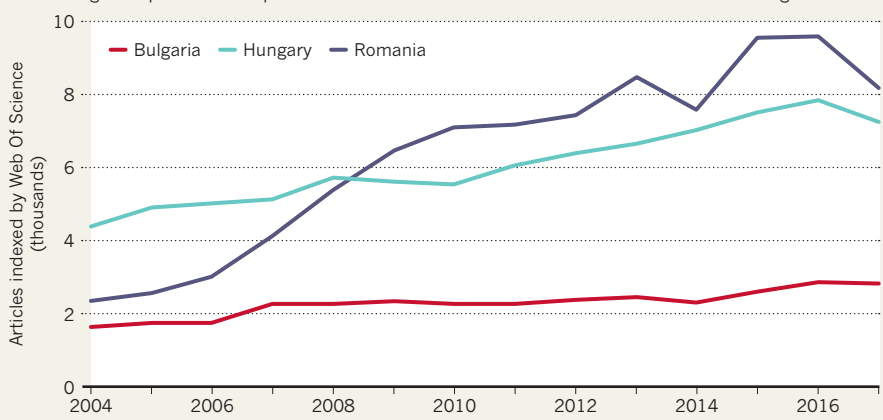
These plans are now at risk, unless Bulgaria can persuade the EU’s regional-policy directorate general to release the frozen funds. Meanwhile, the 2018 science and higher education budget stands at 2013 levels: just 415 million leva (US\$263 million), plus another 98 million leva for the Bulgarian Academy of Sciences.

The financial difficulties also threaten Bulgaria’s national research-infrastructure road map, which was published in June 2017. Kostadin Kostadinov, an adviser to the country’s science and education minister, Krasimir Valchev, says that the road map “will increase research potential in Bulgaria according to the needs of local industry and regional development”, and that it is part of a plan ultimately to raise the country’s total science spending to 1.5% of gross domestic product (GDP). That figure currently stands at 0.96% of GDP, which is less than half the EU average.

Problems with science funding are exacerbated by corruption, say several scientists. Not only is Bulgaria the poorest country in the EU, it is also the most corrupt, according

### BULGARIA'S OUTPUT LAGS BEHIND

Efforts to boost science in the European Union’s poorest country have been undermined by a lack of funds. Bulgaria’s production of published research has risen at a slower rate than that of its neighbours.



SOURCE: WEB OF SCIENCE

to Berlin-based lobby group Transparency International. Proykova says that science is rarely directly affected by monetary fraud, but corruption makes itself felt in procurement. “For example, things are never delivered to the lab, even though the money has been transferred,” she says. “Or, you get less good equipment for the same money, because the company takes some of the funds.”

Some scientists see Bulgaria’s turn in the EU presidency as a chance for change. Lidia

Borrell-Damián, director for research and innovation at the European University Association in Brussels, says that it provides an opportunity for Bulgaria’s universities to connect with others. Daniel Smilov, a political scientist at Sofia University, hopes that the presidency will put the country’s problems on the map, forcing change from outside that has been lacking from within. “It is an important moment,” he says, “because our visibility will be great.” ■

### DNA SEQUENCING

# Super-invasive crayfish revealed to be a genetic hybrid

Scientists examine DNA of a marbled crayfish that is spreading ferociously.

EWEN CALLAWAY

Molecular biologists have sequenced the genome of an invasive species of crayfish that can reproduce without mating and is spreading rapidly across Madagascar. The marbled crayfish (*Procambarus virginalis*) was first spotted in aquariums in Germany in the 1990s. Now, DNA sequencing suggests that the species is probably the product of two distantly related members of a different crayfish species, the team reported on 5 February in *Nature Ecology and Evolution*<sup>1</sup>.

The marbled crayfish has already been banned in the European Union and some parts of the United States because of the threat it poses to freshwater ecosystems. The species has now spread into the interior of Madagascar and risks crowding out seven native crayfish

species. “This is a very aggressive population,” says Frank Lyko, a molecular biologist at the German Cancer Research Center in Heidelberg, who co-led the study. “If the marbled crayfish continues to explode at its current pace, it will probably outcompete endemic species.”

**“If the marbled crayfish continues to explode at its current pace, it will probably outcompete endemic species.”**

The marbled crayfish carries three copies of each chromosome, instead of the usual two<sup>2</sup>. Lyko and his team sequenced the genome of a single individual from a laboratory strain known as Petshop. Its DNA revealed a surprise: it had two different genotypes at many places in its genome. The best explanation for this pattern, says Lyko, is that two of the chromosomes

are nearly identical in sequence, but the third differs substantially.

The two distinct genomes are closely related to those of another freshwater crayfish, *Procambarus fallax*, native to Florida and popular with aquarists. Lyko speculates that marbled crayfish emerged when the genome of a sperm or egg of one *P. fallax* individual became duplicated, which can happen in response to sudden changes in temperature. If this cell was then fertilized by another individual living in the same aquarium, it would have resulted in an embryo with three copies of its genome, says Lyko. This would represent a new species. Lyko says that the first marbled crayfish was probably born in an aquarium in either Germany or the United States, and its offspring widely shared between fish collectors.

The first scientific description of the marbled crayfish appeared in 2003, in a ▶

**CORRECTION**

The News story 'Super-invasive crayfish revealed to be a genetic hybrid' (*Nature* **554**, 157–158; 2018) incorrectly stated that Julie Jones was the first to identify marbled crayfish in Madagascar. In fact, another team made the discovery; Jones and her team were the first to survey its spread in the nation.



The marbled crayfish threatens to crowd out seven native species in Madagascar.

► *Nature* paper<sup>3</sup> showing that all members of the species they surveyed were female and reproduced through parthenogenesis — a process by which an unfertilized egg develops into an adult with a genome identical to its mother's. How the first marbled crayfish gained the ability to reproduce through parthenogenesis is a mystery, says Lyko.

To better understand the species' spread,

Lyko's team did more-limited DNA sequencing of 49 individuals caught across Madagascar. These studies showed a stunning lack of genetic diversity, owing presumably to the species' recent origin and ability to reproduce through parthenogenesis.

Julia Jones, a conservation scientist at Bangor University, UK, who first identified<sup>4</sup> marbled crayfish in Madagascar in 2007, says that the

species' spread is due largely to their popularity as a food source. In 2009, she met a man on a bus carrying a plastic bag full of them that he planned to dump into his rice fields in the hope of creating a sustainable stock, she says.

Stopping their spread in Madagascar will be "almost impossible", says Lyko. Collaborators there have begun campaigns urging people not to transport the creatures or release them into rice fields. The message is a hard sell in a country where poverty levels are high and marbled crayfish are a cheap and popular source of protein. Lyko's colleague brought a few dozen that she had caught to a family barbecue. "This went down quite well," he says. ■

1. Gutekunst, J. *et al. Nature Ecol. Evol.* <http://dx.doi.org/10.1038/s41559-018-0467-9> (2018).
2. Martin, P., Thonagel, S. & Scholtz, G. *J. Zool. Syst. Evol. Res.* **54**, 13–21 (2016).
3. Scholtz, G. *et al. Nature* **421**, 806 (2003).
4. Jones, J. P. G. *et al. Biol. Invasions* **11**, 1475–1482 (2009).

#### CORRECTION

The News Feature 'The science that's never been cited' (*Nature* **552**, 162–164; 2017) originally included a link to the data behind the charts. *Nature* has subsequently been told that the data are not available to make public, so the link has been removed online.



Wanga Zembe-Mkabile felt like both an insider and an outsider doing community research.

# SOUTH AFRICAN SCIENCE FACES ITS FUTURE

**W**anga Zembe-Mkabile learnt a lot about herself from being uncomfortable in other people's kitchens. In 2009, the South African social-policy researcher was collecting data for her PhD on the outcomes of government child-support grants. The research called for 'cupboard inventories' — taking stock of the food in study participants' kitchens. But seeing the embarrassment in home after home as people opened their often-empty pantries, Zembe-Mkabile felt something was amiss. "It just didn't feel right to look into people's cupboards," she says.

At the time, she did not act on her unease. Only years later, as an established scientist, did Zembe-Mkabile begin to understand the complexity of her apprehension. Community-based research often puts young scientists in a position of power over research participants, a role that can be daunting and unfamiliar. But for Zembe-Mkabile, the feelings went deeper. She'd known apartheid, and how its architects had used science to underpin their racist philosophies. The vestiges of that power imbalance were still there in her kitchen encounters.

Zembe-Mkabile grew up around poverty, but as a scientist trained at the University of Oxford, UK, she was the product of a system shaped by and for white Europeans. The tension between these two roles — the ►

*Amid a tumultuous political landscape, a generation of black researchers is gearing up to transform South African science.*

BY LINDA  
NORDLING

► insider and the outsider — is central to her identity as a researcher, and has shaped her thinking about how research can, and should, be done. Now, working at the South African Medical Research Council in Cape Town, she directs studies on how social policy relates to poverty, inequality and health. She plans to involve communities at the design stage of her experiments and, eventually, to include them in the analysis as well. Already, if she feels a tool or question is not appropriate for its setting, she eliminates it from her research. “Some questions are not worth exploring if they are going to trample on people’s dignity,” she says.

Zembe-Mkabile thinks about her experiences a lot when she considers the mounting calls in her country to decolonize academia. Decolonization is a movement to eliminate, or at least mitigate, the disproportionate legacy of white European thought and culture in education. According to advocates, this is not just about increasing the number of black scientists, although such racial ‘transformation’ is an important part of the process. It also means dismantling the hegemony of European values and making way for the local philosophy and traditions that colonists had cast aside. Substantial literature from around the world supports the need to change curricula, and some South African universities have begun to take action and establish review committees. But the push for change is sometimes tense. Student demonstrations have wrapped arguments about decolonization into protests over university fees, and have resulted in disrupted classes, fires and millions of dollars spent on security and repairs.

Science departments have struggled to define what decolonization means for their curricula and for research. Most are ramping up efforts to overcome the glaring under-representation of black scientists, but what comes next is unclear. Zembe-Mkabile’s generation, which straddles pre- and post-apartheid South Africa, will soon be leading the country’s research institutions as they grapple with the challenge of reformulating science for the new South Africa.

## THE ROLE MODELS

South Africa, like many nations, is currently dealing with high unemployment rates and glaring inequality. These are cast into sharp focus by the legacy of apartheid rule. Although political power has been in the hands of the black majority since the dawn of South African democracy 24 years ago, economic power remains with white people: white households in 2015 earned around 4.5 times as much as black households, and whites hold more than 60% of top management positions, despite accounting for only 10% of the working population. In universities, black people account for not quite 35% of academics, despite making up about 80% of the population. Students, meanwhile, face multiple barriers to achievement, including an education system that has left many unprepared for university studies. A 2015 government report found that black South Africans had the highest dropout rate in the country; 32% leave their studies in their first year. As for curricula, African literature, philosophy, medicine and culture are often relegated to optional courses or skipped entirely.

It is against this backdrop that researchers in Zembe-Mkabile’s generation forged their academic paths. Children during apartheid, they reached adulthood in the rainbow-coloured afterglow of Nelson Mandela’s release from prison in 1990. Some hail from communities that are distrustful of science. In Xhosa, Zembe-Mkabile’s home language, there isn’t even a word for research. The best approximation, she says, is *ukuphanda*, which has negative connotations. “It means to search for a bad thing, like a police investigation,” she says.

The scientists of Zembe-Mkabile’s generation are role models for the generation born after 1994, known locally as ‘born frees’. According to a study<sup>1</sup> published last year, this generation is predicted to boost the country’s proportion of black researchers to more than 50% by 2025. That’s a heavy burden for those like Zembe-Mkabile squeezed between

the demands of the academic system that trained them and the expectations of a youth clamouring for radical change. Zembe-Mkabile says that protesting against inequality in the universities was not on the table when she was a student. “You entered these spaces and you were so grateful to be there that you didn’t question anything. We were fast asleep. At least now, students are alert.”

## DECOLONIZING THE MIND

Zembe-Mkabile’s experience is not unique. Amanda Hlengwa, an academic developer at Rhodes University in Grahamstown, has similar memories of her undergraduate degree in Durban in the late 1990s. “The goal was assimilation. That was the only way to survive.” This is changing, she says: universities are beginning to recognize students’ diverse backgrounds, and the challenges that university culture presents. But strategies to address this gap have been slow to materialize and are unevenly implemented.

Thaddeus Metz, a philosopher at the University of Johannesburg, agrees. A white American who settled in South Africa in 2004, he was the first to teach African philosophy at the nearby University of the Witwatersrand, the city’s most prestigious research university, where he worked before his current post. “There is this long-standing intellectual tradition that has been neglected at best, at worst denigrated,” he says. He adds that the majority of students, regardless of their race, are curious about African knowledge traditions, but that there’s a lack of institutional leadership. Many in the humanities and social sciences are angry because they feel isolated and powerless.

In the natural sciences it gets more complicated, because the meaning of decolonization is not well defined and its relevance is contested. Does decolonizing science mean throwing out Isaac Newton, Charles Darwin and Gregor Mendel, and starting afresh with indigenous knowledge? Such demands have been made, most famously by a University of Cape Town student in an online video of a campus discussion titled ‘Science must fall?’. Metz says he’s encountered the argument. “Some of my colleagues think that if something hasn’t come from Africa, it’s somehow disqualified.”

But only a small minority of scientists hold such radical views. For most, decolonization of science calls for something more complex and subtle. “Decolonization is going to happen in the mind,” says Siyanda Makaula, a former cardiology lecturer who now works in university governance. Such shifts in thinking could mean, for example, that pharmacology students hear how drugs are being developed from plants their grandmothers used to treat stomach ache. This would show the relevance of traditional culture in modern science and anchor the curriculum in local experience. In other subjects, it could be about highlighting the contribution of non-Europeans, or facing the unsavoury history of a discipline: for example, exploring how medical research had a role in fuelling racist ideas and how these were challenged and overturned. Across the board, it means ensuring that research addresses local problems and challenges.

**Nokwanda Makunga**  
was warned off joining  
the faculty of a formerly  
all-white university.



**“THE GOAL WAS ASSIMILATION.  
THAT WAS THE ONLY WAY TO SURVIVE.”**



Makaula thinks that scientists often hide behind their disciplines' putative universality — that a cell is a cell, whether it belongs to an African or a European, or that the laws of physics apply to all — to avoid the need to question the way they do things. "It's an excuse they use," he says. But the point of science, he adds, is to find solutions for real-world problems. And for that, context needs to be part of how science is taught, he says. "It's about how you teach it, how you apply it, how you make it relevant, so the person can receive it and absorb it better."

Such refocusing is taking much too long in South African universities, says Makaula. And that inertia is costing the country dearly in terms of black research talent. He sees himself as a prime example. A decade ago, Makaula earned a PhD in cardiology. But repeated brushes with racism and tokenism — being asked, along with other black students, to meet potential funders while his white colleagues could stay in the lab — frustrated him to the point that he left academia. Today, he works for the Council on Higher Education based in Pretoria, a public-sector body that deals with quality control and regulatory compliance in universities.

On the face of it, South African universities are working on decolonizing their academic offerings. Most have created committees to review their curricula — although few have much to show for it. And all are under pressure from government and funding bodies to train and hire more black academics. Research funders are following suit. A few years ago, the Medical Research Council dedicated a significant portion of its largest grant programme to early-career scientists, and added weighting for gender and race. The proportion of the grants going to white investigators has since shrunk, from 72% in 2012 to 37% in 2016. The council is also working on a position statement on decolonization to sharpen its efforts to recruit black scientists, says Glenda Gray, the council's president (see

"Three cultures"). It will look at how medical research can draw on social science to become more sensitive to community needs. "You only get true well-being if you understand the context in which the biological happens."

#### 'NO PLACE FOR A BLACK WOMAN'

Some South Africans approach decolonization as a way to rediscover their heritage. Nokwanda Makunga, a biotechnologist at Stellenbosch University near Cape Town, grew up in the intellectual circles that gave rise to anti-apartheid freedom fighters such as Steve Biko and Nelson Mandela. From a young age, Makunga knew exactly what a scientist did. One of her early memories is of helping her father — a botanist — count kernels of maize (corn) for an experiment. In the dying years of apartheid, she attended a private boarding school in Grahamstown, where racial tensions were muted. It therefore came as a shock when, in 1990, she arrived at university in Pietermaritzburg in the politically fractious province now known as KwaZulu-Natal. "I came from a bubble that was non-racial, non-political. Then I was launched into the true South Africa." It was a struggle. Her superior education and clipped private-school vowels singled her out as "too white" to belong with the black students. But she was also too black for the white students. "I was getting it from both ends."

In 2004, after earning her doctorate, Makunga yearned to move to a quieter, more research-focused institution. She got an offer from Stellenbosch University, a formerly all-white institution nestled in the picturesque Cape Winelands. It offered stability, and a platform for Makunga to build an international reputation. But it had also, historically, been a bastion of white supremacy, having produced infamous apartheid-era prime ministers such as Hendrik Verwoerd and D. F. Malan. Some of Makunga's friends were horrified. "One told me that Stellenbosch is no

## THREE CULTURES

## FOREIGN BLACK RESEARCHERS FACE ADDITIONAL CHALLENGES IN SOUTH AFRICAN ACADEMIA

Black researchers are rapidly moving into South Africa's academic spaces. But not all of them are considered 'black' by the country's Department of Higher Education and Training. Researchers from other parts of the world are instead classified as 'foreign'.

It's a large and fast-growing segment. One report<sup>4</sup> found that although black PhD graduates outnumbered whites for the first time in South Africa's history in 2012, more than half of them hailed from countries such as Nigeria, Zimbabwe, Uganda and Kenya.

There are a few reasons for this. It is cheaper to study in South Africa than in Europe or the United States, and the country offers better research facilities than elsewhere in Africa. But for some locals, the growing presence of foreign black researchers is a problem. Jobs are scarce, and some believe that universities are more willing to hire non-South African black people than locals.

So foreign black scientists — such as Thumbi Ndung'u, a Kenyan virologist based at the University of KwaZulu-Natal in Durban — experience a special kind of alienation. "You can't completely identify with the local black population. They see you as an outsider. On the other hand, you are not in the white old-boys' club," he says.

Ndung'u had anticipated some friction when he moved to Durban in 2005 to study HIV. But it wasn't until he lived there that he began to understand the frustrations of local black academics. The system is blind to its own biases, he says.

Most of Ndung'u's own graduate students are black South Africans. They face many challenges, he says, but, given the right support, they blossom. "There needs to be express effort to get them into the system. So South African universities don't continue to have this problem in the future." **L.N.**

place for a black woman. He didn't say why, just that it was very conservative." Makunga took it as a challenge. "How will it ever be a place for black women if no black women are willing to go there?"

Makunga's research has brought her closer to her roots. When she was growing up, her family did not use traditional medicine, she says. But now she studies South African medicinal plants, using modern biotechnology to explore their pharmacological properties — and she reckons the work is "pretty decolonial". Having studied a variety of plants, in 2016 she returned to the Eastern Cape, where she grew up, to learn about the traditional medicine practised by her ancestors. She takes her responsibility as custodian of these practices seriously. "I'm holding somebody else's knowledge. I need to treat it with respect," she says.

Stellenbosch has changed a lot since she was warned off, says Makunga, who is currently at the University of Minnesota in Minneapolis for a nine-month Fulbright scholarship. She feels welcome at Stellenbosch and valued as a black woman — still a rare occurrence at faculty level. Still, she longs for the day when that is not the headline issue; when she can be a scientist first and a black woman second. "I would like us to move beyond our apartheid race hangover," she says wistfully.

Black women are among the most under-represented groups in South Africa's academic melting pot. They make up 14% of the country's researchers, compared with black men's 18%. And they face tough odds. In her 2015 article 'Leadership: The invisibility of African women and the masculinity of power', Mamokgethi Phakeng writes<sup>2</sup> that black women, as well as being marginalized for their gender and race by white society,

face opposition from patriarchal African cultures. This "masculinity of power", she writes, needs to be challenged alongside colonialism and sexism.

Phakeng is a scholar of mathematics education and deputy vice-chancellor of research at the University of Cape Town. Her outspokenness about the experience of black researchers has charmed an army of Instagram and Twitter followers. But at times the veneration is misguided, she feels. It's great to inspire young people to speak up and be themselves, "but I don't want that to be my most powerful role", she says.

To her, the most important thing she has done is excel as a researcher. Phakeng's work centres on mathematics and language. She showed<sup>3</sup>, for example, that code-switching — alternating between languages — helps multilingual people to understand mathematical concepts. This is significant in South Africa, where students have been scolded for using their home language. Today, code-switching is encouraged in many classrooms.

Yet, critics have argued that her field is not a suitable background for the head of research at one of Africa's strongest science institutions. Last October, an e-mail started circulating, questioning her qualifications. She took on her attackers, and vice-chancellor Max Price denounced the e-mail and its contents.

The future of South Africa's university sector is uncertain. Hlengwa worries that the momentum created by student protests might fizzle out without sustained change taking root.

"While the heat was on, you had opportunities to work on transforming curricula," she says. But as universities learn to work with unrest, they snap back to old ways, she says. She also worries that black academics are being run into the ground by the demands placed on them — from being called on to sit on diversity committees to giving advice on the complex challenges facing black students and staff. "Where's the space for me to do some deep thinking about my research?" Hlengwa asks.

It is a burden and a challenge. And Phakeng argues that it can be helped only by discourse. One of the things she has done since joining the University of Cape Town in mid-2016 is to speak to its black South African academics. For some, she says, it's the first time they've been called on by management to share their experiences. "I ask people, what stories do you tell yourself? Those stories shape the possibilities of what we can do." ■

**Linda Nordling** is a freelance journalist in Cape Town, South Africa.

1. Breetzke, G. D. & Hedding, D. W. *Higher Educ.* <http://dx.doi.org/10.1007/s10734-017-0203-4> (2017).
2. Phakeng, M. S. *Afr. J. Sci.* **111**, a0126 (2015).
3. Setati, M., Adler, J., Reed, Y. & Bapoo, A. *Lang. Educ.* **16**, 128–149 (2002).
4. Cloete, N., Sheppard, C. & Bailey, T. in *Knowledge Production and Contradictory Functions in African Higher Education* (eds Cloete, N., Maassen, P. & Bailey, T.) Ch. 5 (African Minds, 2015).



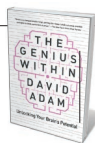
**Mamokgethi Phakeng** is ambivalent about being a role model for young black academics.

ALEXIA WEBSTER FOR NATURE

# COMMENT

**EARTH SCIENCE** A geologist looks for answers among the oldest rocks on Earth **p.166**

**PSYCHOLOGY** A personal journey to the wilds of brain boosting **p.168**



**TECHNOLOGY** Bitcoin has the carbon footprint of a small nation **p.169**

**ZOOLOGY** Baleen species is on the brink of extinction for the first time in 300 years **p.169**

WATERFRAME/ALAMY



A floating forest of *Sargassum* seaweed in the Sargasso Sea provides food and shelter for hatchling loggerhead sea turtles in the Atlantic Ocean.

## Protect the neglected half of our blue planet

Maintaining momentum is crucial as nations build a treaty to safeguard the high seas, argue **Glen Wright, Julien Rochette, Kristina M. Gjerde and Lisa A. Levin.**

At the close of 2017, 14 million UK viewers tuned into the acclaimed second series of David Attenborough's *Blue Planet*, making it the year's most-watched television show. It brought the wonders of the ocean into people's living rooms and captured the public imagination as never before. Now is the time to capitalize on this enthusiasm, and to advocate for strong, legally binding protections for the high seas — the almost two-thirds of our planet's ocean that are beyond the control of any one state (see 'Neglected waters').

A start has been made. A landmark resolution was adopted at the United Nations General Assembly on 24 December 2017, marking the beginning of formal diplomatic negotiations for an international treaty to conserve and sustainably use the high seas. Co-sponsored by more than 130 nations, Resolution 72/249 is the result of more than a decade of scientific debates, legal controversies and political wrangling<sup>1</sup>. The decision paves the way for a range of measures, including a much-needed system of global marine protected areas (MPAs) to sustain

aquatic life in a rapidly changing ocean.

Now we must ensure that real progress is made over the next few years. The treaty, expected some time after 2020, will need to include provisions for firm international oversight and direction if it is to have any chance of overcoming problems with the existing regulatory framework. As with any such negotiations, there is a risk that they will result in a toothless call for 'urgent' action and increased cooperation.

Non-governmental organizations and environmental groups will continue to ►

► push for strong conservation measures, including strictly protected reserves. The research community can contribute by speeding up the collection of baseline data to characterize existing environments and biodiversity; coordinating observation efforts across disciplines; monitoring and assessing ocean health; and further studying how MPAs and other conservation tools work on the high seas. Indeed, ocean science could be a unifying focus for this new agreement<sup>2</sup>. Social scientists, legal scholars and other experts can feed the negotiations with pragmatic options.

Together, we must advocate for a strong international treaty if crucial high-seas ecosystems are to survive and thrive.

### PATCHY PROTECTION

Governments have repeatedly made high-level political commitments to conserve marine biodiversity. The Aichi Biodiversity Targets and the UN Sustainable Development Goals, for example, demand protection of 10% of the world's ocean (although some scientists argue that at least 30% is necessary<sup>3</sup>). This is to preserve wild spaces, sustain fisheries, protect the ecosystems that regulate the climate and preserve a wealth of biodiversity<sup>4</sup>. Governments have nonetheless been slow to act. Just 4% of the ocean is currently protected, and hardly any MPAs cover the high seas<sup>5</sup>.

Marine areas beyond national jurisdiction are regulated by a patchwork of different agreements and institutions, each with their own peculiarities and pitfalls. Most of these organizations focus on the management of a particular resource or activity. The International Seabed Authority (ISA) oversees seabed mining. Regional fisheries-management organizations regulate high-seas fisheries. And the International Maritime Organization sets out shipping rules.

There are few channels of communication between these agencies, much less formal cooperation or coherence between their management measures. Their decisions are highly politicized, and the need to reach consensus among member countries can trump scientific evidence.

Some regional initiatives have made limited progress. The OSPAR Commission, named after its original conventions in Oslo and Paris, is composed of 15 countries and the European Union. It has designated ten MPAs in the high seas of the northeast Atlantic. However, these apply only to its member countries, and OSPAR does not have the authority to regulate many activities or to ensure that conservation is part of fisheries decisions.

In 2017, the ISA approved a 15-year exploration contract with Poland, covering part of the Mid-Atlantic Ridge. Within this OSPAR area sits the Lost City hydrothermal

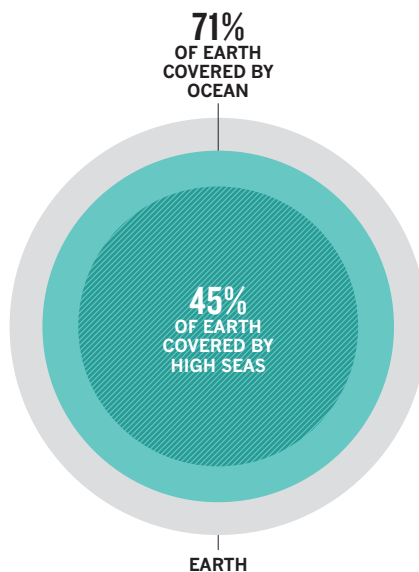
field, a unique range of 60-metre-tall calcium carbonate chimneys. The UN Educational, Scientific and Cultural Organization (UNESCO) and the International Union for Conservation of Nature (IUCN) have highlighted that the site might meet the criteria for World Heritage status<sup>6</sup>. The ISA did not consult UNESCO, the IUCN or OSPAR, which left scientists who study this feature with no avenue for input other than writing a letter of concern after the contract was approved.

On the other side of the Atlantic, the Sargasso Sea Commission is attempting to protect a unique floating forest of *Sargassum* seaweed — recognized as an Ecologically or Biologically Significant Marine Area (EBSA) under the Convention on Biological Diversity. This recognition does not entail any management measures, and only one of the few regulatory bodies in this high-seas region has shown an interest in implementing any.

In the Southern Ocean, countries have worked together within the dedicated Antarctic Treaty System to designate part of the Ross Sea as the world's largest MPA (1,549,000 square kilometres). This required

### NEGLECTED WATERS

Almost two-thirds of the planet's ocean are classed as international waters or high seas, meaning that they are beyond the control of any one state.



intense diplomatic efforts. Nonetheless, the resulting protections are limited, and the process to establish more MPAs recently stalled.

### CONSERVE AND CONNECT

Research has shown that MPAs are effective if they are done right. Large, long-term, 'no-take' reserves that are isolated by deep water or sand and backed up with strong enforcement have five times more large-fish biomass than unprotected areas<sup>7</sup>.

Recent advances have greatly improved the evidence base for MPAs on the high seas, dispelling many common assumptions about their feasibility and efficacy<sup>8</sup>. For example, scientists previously thought that species ranges were too big to designate meaningful MPAs, but we now know that even wide-ranging deep-water species assemble to feed and spawn, and use particular habitats for nurseries<sup>9</sup>. So strategically protecting just part of a species' range could help to sustain populations<sup>10,11</sup>.

It is easy to find candidates for an initial suite of MPAs. UNESCO has identified 5 possible high-seas World Heritage Sites; nearly 50 EBSAs cover portions of the high seas; fisheries bodies, following requirements in UN resolutions, have identified 'vulnerable marine ecosystems' susceptible to impacts from bottom trawling; and the ISA is identifying 'areas of particular environmental interest'. These designations cover a broad range of habitats, from deep-water coral grounds to abyssal plains, and are grounded in scientific criteria — including a site's uniqueness, productivity, complexity and fragility.

Protecting such sites is a start, but will not insure the ocean against the many threats it faces. A wider network of representative and connected MPAs will be needed to provide resilience to climate change and to maintain biodiversity by ensuring links between migration routes and spawning grounds<sup>12</sup>. No one has worked out where, how large or how deep these areas should be. Complementary measures and better management will also be needed for the ecosystems and activities that fall outside MPAs.

Calling for protection of a swathe of the high seas might seem starry-eyed. But some have made the case for entirely closing the high seas to fishing, arguing that this would lead to greatly increased fisheries yields and profits overall<sup>13</sup>.

More research will be required if we are to protect these deep and distant seas effectively. Ramping up basic research efforts to improve baseline data is crucial, as is improving our understanding of how climate change and other stressors affect invaluable ecosystem services. We will need to better coordinate and expand existing observation programmes, improve data access and promote training for young scientists. Next-generation molecular, computing, telemetry and observing technologies must also be developed and applied.

Some progress is being made here: researchers are developing techniques for growing deep-sea organisms in the laboratory, shedding light on their reproductive traits and biology. The UN has declared 2021–30 as the Decade of Ocean Science for Sustainable Development. This should help to mobilize the scientific community around these issues.



Great white sharks (*Carcharodon carcharias*) often visit an area of the Pacific Ocean dubbed the White Shark Café. It is one of five areas proposed as a high-seas World Heritage Site.

DAVID FLEETHAM/BARCROFT/GETTY

The questions of who will designate MPAs and how management measures will be implemented are politically charged. The strongest possible outcome of the upcoming treaty negotiations, from a legal perspective, would be a new UN body with wide powers to make binding, top-down decisions, working in concert with existing organizations. At the other end of the spectrum, states could be left to cobble together MPAs within the existing system, with the new agreement providing some form of obligation and oversight. The former would provide a powerful means of protecting this important global commons; the latter might leave conservation beholden to the failings of the current framework, with states likely to continue dragging their feet.

A balance will need to be found. To be effective, the new instrument must provide sufficient international oversight, while respecting the mandates of existing organizations and ensuring that a majority of states is prepared to sign up.

What is certain is that individual states will remain responsible for controlling ships flying their flag. Proactive states could therefore agree to work collectively through the new treaty to protect priority places by controlling the biodiversity impacts of their vessels, while encouraging non-parties to do the same. However, the negotiations are not intended to address 'flags of convenience', whereby a country registers vessels on a 'no questions asked' basis, generally in exchange for a fee.

## GENETIC GOLDMINE

Marine protection is only one part of the treaty discussions. The question of how to regulate the exploitation of marine genetic resources also promises to be both technically and politically challenging.

Genes extracted from marine creatures in the high seas are being used to develop new pharmaceuticals and cosmetics. There is currently no requirement to share the profits that arise from the exploitation of this common resource.

The few states that have the capacity to conduct bioprospecting are keen to maintain the status quo, which is essentially first come, first served. Others want to create a formal

*“Turning good intentions into an effective treaty will take time, money and scientific input.”*

mechanism for sharing the profits, similar to a system already being put in place for seabed mining. The long and complicated chain of discovery makes it difficult to capture any monetary benefits. And the distinction between bioprospecting and 'pure' scientific research, which is permitted by the UN Convention on the Law of the Sea, is often far from clear.

Researchers can help here, too, such as by improving open-access protocols for data and samples. Even in the absence of a comprehensive regulatory framework or an obligation to share profits, a new treaty could still include helpful provisions that would

promote international science cooperation, capacity building and the development and transfer of marine technology<sup>14</sup>.

Turning good intentions into an effective treaty with meaningful MPAs will take time, money and scientific input. There is uncertainty regarding the position and role of the United States, which has not ratified the UN Convention on the Law of the Sea. Developing countries are calling for greater assistance, which will require developed nations to commit considerable resources. But with the majority of countries now in favour of a new agreement, momentum is building.

Tough diplomatic negotiations might nonetheless be necessary to reach consensus on the finer details of the new treaty. The beauty and value of our ocean could be lost all too easily in the windowless halls of the UN's New York headquarters, obfuscated by realpolitik and the arcane details of international law. Political leaders will need to see strong science and public support if they are to develop an ambitious agreement to finally protect the neglected half of our blue planet. ■

**Glen Wright** is a research fellow at the Institute for Sustainable Development and International Relations (IDDRI) in Paris, France. **Julien Rochette** is director of the oceans programme at IDDRI. **Kristina M. Gjerde** is senior high-seas adviser to the International Union for Conservation of Nature in Gland, Switzerland. **Lisa A. Levin** is distinguished professor of biological oceanography at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA. e-mail: glen.wright@iddri.org

1. Wright, G., Rochette, J., Druel, E. & Gjerde, K. *The Long and Winding Road Continues: Towards a New Agreement on High Seas Governance* (IDDRI, 2016).
2. Harden-Davies, H. *ICES J. Mar. Sci.* **75**, 426–434 (2018).
3. O'Leary, B. *et al. Conserv. Lett.* **9**, 398–404 (2016).
4. Rogers, A. D., Sumaila, U. R., Hussain, S. S. & Baulcomb, C. *The High Seas and Us: Understanding the Value of High-Seas Ecosystems* (Global Ocean Commission, 2014).
5. Worm, B. *Nature* **543**, 630–631 (2017).
6. UNESCO & IUCN. *World Heritage in the High Seas: An Idea Whose Time Has Come*. World Heritage Reports 44 (UNESCO, 2016); available at <http://whc.unesco.org/en/highseas>
7. Edgar, G. J. *et al. Nature* **506**, 216–220 (2014).
8. Ceccarelli, D. M. & Fernandes, L. *The Value Of Offshore Marine Protected Areas For Open Ocean Habitats And Species. Report To The MACBIO Project* (GIZ, IUCN & SPREP, 2017).
9. Scales, K. L. *et al. J. Appl. Ecol.* **51**, 1575–1583 (2014).
10. Grüss, A., Robinson, J., Heppell, S. S., Heppell, S. A. & Semmens, B. X. *ICES J. Mar. Sci.* **71**, 1515–1534 (2014).
11. Boerder, K., Bryndum-Buchholz, A. & Worm, B. *Mar. Ecol. Progr. Ser.* **585**, 1–15 (2017).
12. Roberts, C. M. *et al. Proc. Natl Acad. Sci. USA* **114**, 6167–6175 (2017).
13. White, C. & Costello, C. *PLoS Biol.* **12**, e1001826 (2014).
14. Harden-Davies, H. *Deep Sea Res. II Top. Stud. Oceanogr.* **137**, 504–513 (2017).



FERNANDO MOLERES/PANOS

The west coast of Greenland is fertile ground for geologists.

## GEOLOGY

# Epiphanies of the edgelands

**Ted Nield** admires a geologist's layered exploration of Greenland's remote fringe.

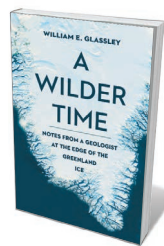
**G**eologist and former surf dude William Glassley has spent six field seasons studying the ancient rock of coastal Greenland. As he probes our planet's youth, three billion years ago, many will envy him. His brief but ambitious *A Wilder Time* demonstrates that there's nothing like geology for acquainting you with the joys of remote isolation at other people's expense.

The area he explores, with Danish colleagues Kai Sørensen and John Korstgård, is vast: part of the coastal fringe of ice-smoothed rock and periglacial tundra that extends like a valance around Greenland's enormous central ice cap. There is sea to the west, crumbling ice cliffs 150 kilometres and more to the east. *A Wilder Time* sees our heroes marooned in this wilderness, alone in the short summer's perpetual day. Glassley eloquently evokes a place where land feathers into Arctic sea, ice floes glide by on mirror-smooth tongues of clear, frigid water and silence reigns.

What drew the companions there might sound, by contrast, like a storm in an academic teacup. Someone (tactfully

left unnamed) had published a paper attacking the established geological view that the study area — between Nordre Isortoq in the south and Disko Island to the north — is part of the roots of an ancient mountain range, the Nagssugtoqidian mobile belt. Geologists are familiar with these Inuit place-names, many ending in 'oq'. Pronunciation should sound, the late Stephen Moorbath (an isotope geochemist and geochronologist) once told me, "like a piano string being cut at the bottom of the ocean".

Moorbath helped to make the area famous by finding what are still among the oldest known rocks on Earth, almost 3.8 billion years old. In the 1960s and 1970s, geologists



**A Wilder Time:**  
Notes from a  
Geologist at  
the Edge of the  
Greenland Ice  
WILLIAM E. GLASSLEY  
Bellevue Literary Press:  
2018.

Arthur Escher and Juan Watterson mapped these high-grade metamorphic melanges of altered sediments, mantle rocks and ocean-floor basalts. In the 1980s, Feiko Kalsbeek, Bob Pidgeon and Paul Taylor interpreted it all in the light of plate tectonics. The distinctive east-west shear zones that transected the region were, they said, sutures left by the most ancient plate-tectonic collisions on our planet, during the early Proterozoic eon, which began around 2.5 billion years ago. But seeds of doubt were cast. The new paper made fundamental challenges to earlier interpretations that seemed themselves so misguided and riddled with errors and misconceptions that they could not go unanswered.

Such is the scientific narrative underpinning *A Wilder Time*, whose rather over-complicated structure arrives at a satisfying conclusion. The epilogue demonstrates how Glassley's team confirmed and even refined the original interpretation of the mobile belts, putting its assailants to flight.

This story offers perspectives on deep time to boggle minds, from the immense ages of

the rocks and events concerned. Metamorphic petrology is no easy material for popular science. By the time you explain how the phyllites, schists and gneisses started life, why they were taken to the depths of Earth's crust and how their minerals were changed under combinations of heat and pressure (each producing distinctive suites of new minerals), many readers will have lost interest. Wisely, Glassley doesn't try too hard — which is fine, because the science is almost a narrative ploy.

Natural scientists may be the only intellectuals these days who find themselves routinely exposed to the transformative experience of wilderness. Yet (as I have seen during desert fieldwork in the Middle East) on many of them it seems wasted. This may not be their fault. Expeditions, such as Glassley's, are a lesson in how travel can narrow the mind. It is hard enough to keep focused on the work when trying to cope with midges, heat, cold, disorientation, altitude and homesickness, never mind dehydration, disgusting camp food and the physical consequences. It takes a deep attunement to the wild's allure to keep appreciating the surroundings.

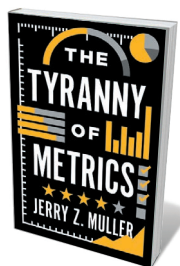
Glassley's vivid impressions of East Greenland attempt what few scientist-writers try: to explore beyond the comfort zone of his field. Followers of the medieval philosopher Duns Scotus coined the terms *haecceity* ('thisness', of specific things) and *quiddity* ('whatness', of the classifier). These are 'science'. Glassley tries also to grasp something beyond: the noumenon, an ineffable inner reality in things that cannot be discerned by the senses.

Not everyone experiences this psychological epiphany. Scientists sometimes have it trained out of them by the relentless insistence on cold measurement. Glassley, by contrast, seems obsessed with our limitations when it comes to grasping the wholeness of the world. He questions, for instance, how our 'reality' contrasts with, say, a seal's, or a fish's. Absenting himself from camp, he wanders alone with his reflections, and attempts closer communion with the hidden genius of place.

Although he repeatedly explains what he's attempting (a scientist's tendency to write abstracts for everything?), he is not always successful; yet I enjoyed and admired the attempt. What he gropes for requires art, not analysis. Perhaps that was why I kept returning to Hugh MacDiarmid's great 1934 poem, 'On a Raised Beach', which explores the limitations of science in expressing the wholeness of nature. After an opening parody of scientific language, the poet observes: "Deep conviction or preference can seldom/Find direct terms in which to express itself". ■

**Ted Nield** is editor of *Geoscientist* and author of *Supercontinent*. In an earlier life, he, too, sensed the noumenon in remote places at other people's expense.  
e-mail: ted.nield@geolsoc.org.uk

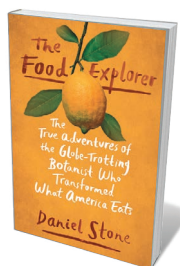
## Books in brief



### The Tyranny of Metrics

Jerry Z. Muller PRINCETON UNIVERSITY PRESS (2018)

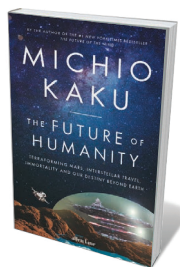
Economic historian Jerry Muller delivers a riposte to bean counters everywhere with this trenchant study of our fixation with performance metrics — a cultural ubiquity saturating education, medicine, finance and governance. As he argues, this reductive approach to monitoring efficiency almost inevitably backfires. It can lower morale by riding roughshod over professionals' experience; invite manipulation, from "gaming the stats" to "teaching to the test"; discourage innovation, promote short-termism; and reward dumb luck. Metrics, he asserts, can usefully bolster judgement, but not supplant it.



### The Food Explorer

Daniel Stone DUTTON (2018)

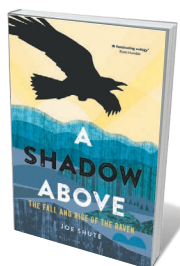
In the annals of intrepid botanists combing the globe for novel species, David Fairchild is a name to conjure with. At the turn of the twentieth century, the plant scientist introduced 200,000 'exotic' species to the United States, then something of a culinary blank slate. Kale, avocados, mangoes, pomegranates and even quinoa began to work their way into US consciousness and, eventually, markets. Daniel Stone's rip-roaring tale takes us from Fairchild's youthful meeting with naturalist Alfred Russel Wallace in Kansas to collecting trips across more than 50 countries, from Trinidad to China.



### The Future of Humanity

Michio Kaku DOUBLEDAY (2018)

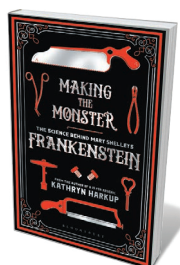
This latest foray into futurism by Michio Kaku sees the physicist unbowed by woes political and planetary. As a master of the long view, Kaku plots humanity's path to becoming a "multiplanet species". He marshals fresh advances in artificial intelligence, nanotechnology and bioengineering for his vision, segueing from lunar stations and Martian colonies to interstellar travel and human genetic engineering. There's plenty of hypothetical innovation, too: ramjet fusion machines, antimatter engines and "laser porting" of human connectomes to enable bodiless exploration of the cosmos.



### A Shadow Above: The Fall and Rise of the Raven

Joe Shute BLOOMSBURY (2018)

Size-wise, the king of corvids is the raven. But for journalist Joe Shute, the bird is also an emblem of our age, caught between the ebb of wilderness and the hope of regeneration. In Britain, after a long, catastrophic decline, numbers have bounced back by 45% over the past two decades. Celebrating that fact, Shute gets inside the skin of the 'feathered ape' with the "rhino-horn beak" and aerial virtuosity. That journey becomes a rich and beguiling tangle of cultural and natural history, birding diary and account of corvid fandom — Charles Dickens being one notable devotee.



### Making the Monster

Kathryn Harkup BLOOMSBURY SIGMA (2018)

Chemist Kathryn Harkup's scientific contextualization of Mary Shelley's *Frankenstein* at 200 is a worthy addition to a crowded shelf. She explicates how trailblazing discoveries in galvanism, chemistry and anatomy helped to form the bones of the book, while its heart beat to the rhythm of Shelley's radical intellectual lineage and milieu. Harkup's handling of Shelley's own story and the literary alchemy wrought by this brilliant teenager compels, not least on how the science fiction has seeped into science fact. [Barbara Kiser](#)

# Into the wilds of brain boosting

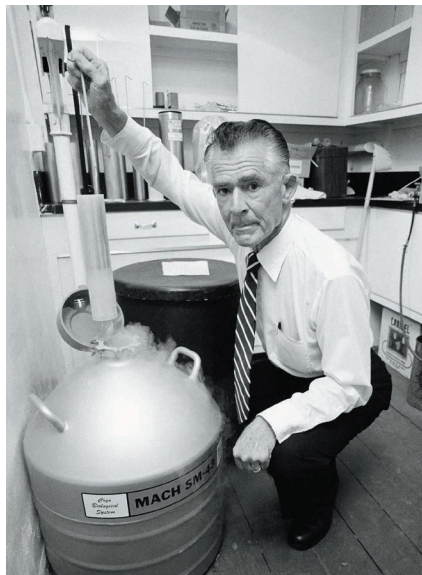
**Trevor Robbins** lauds a personal journey aimed at cracking cognitive enhancement.

Is there a common element that binds diverse mental abilities, from language to mental arithmetic? Or do these skills compete for our brains' limited resources? In *The Genius Within*, David Adam tackles the controversial topic of intelligence, in a follow-up to his exceptional book on obsessive-compulsive disorder (OCD), *The Man Who Couldn't Stop* (Picador, 2014).

Adam (*Nature's* Editorials editor, with whom I worked on a film about the use of animals in OCD research) deftly surveys attempts to test intelligence starting more than a century ago, by statistician and eugenicist Francis Galton and others. Psychologist Charles Spearman introduced *g* as a measure of overall performance across, for instance, verbal and spatial capacities. Alfred Binet introduced the first practical test for intelligence quotient (IQ), purporting to determine a child's 'mental age'. In the ensuing decades, intelligence testing has come under fire for cultural bias, as well as limited scope; it excludes 'emotional intelligence', for example. Adam explores the dissent, from calls for recognition of neurodiversity to the furore over Richard Herrnstein and Charles Murray's 1994 book *The Bell Curve* (Free Press).

Adam's book hinges on one question: can we beef up our intelligence? Ever perspicacious and intrepid, Adam turns guinea pig, exploring means from 'brain training' to cognitive-enhancer drugs. He takes modafinil (licensed for treating the sleep disorder narcolepsy) and finds, consistent with lab studies, enhanced mental focus as he works: "I did feel different ... like I was concentrating on the words I wrote in a more deliberate way." He also endures bouts of crude electrical brain stimulation that simulates the effects of transcranial direct current stimulation (tDCS), administered by his spouse. In the lab, tDCS has been claimed to promote learning by exciting or inhibiting neural activity across specific regions of the cortex.

Adam concludes there is evidence that intelligence can be boosted. Improving cognitive performance is important in people affected by Alzheimer's disease, for instance, or conditions such as attention deficit hyperactivity disorder. But he argues that enhancing 'normal' performance could also be a viable goal. The claim that most people perform optimally is palpably false: think of the effects of fatigue on exam performance. Moreover, surveys reveal that students and academics use cognitive-enhancer drugs.



Robert Klark Graham ran a 'genius sperm bank'.

Yet successful cognitive enhancement carries myriad ethical implications. Is it fair to take modafinil while studying for exams? Could an employer insist on treatment for employees? How would all this be regulated?

Importantly, does cognitive improvement carry a neurological price? Is it possible to enhance *g*, or does boosting one aspect of intelligence, such as working memory, degrade another? A study last year found that modafinil can enhance the performance of people playing chess against computers, but they lose more games through time defaults (A. G. Franke *et al. Eur. Neuropsychopharmacol.* 27, 248–260; 2017).

Adam makes several fascinating digressions into the grisly underbelly of cognitive neuroscience. The Repository for Germinal Choice, for instance, was eugenicist Robert Klark Graham's scheme for selling sperm from Nobel prizewinners. The nineteenth-century Society of Mutual Autopsy sought to advance neuroscience by dissecting dead members' brains. Adam reveals how assessment of death-row prisoners' intellect can halt, or implement, execution. Most remarkable

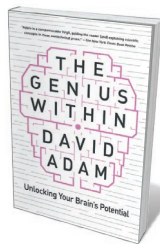
are his discussions of extraordinary mathematical or musical abilities in people with, for instance, autism or brain damage. This sort of neurodiversity highlights a need for clearer understanding of whether specific brain regions actively compensate for loss of function due to damage, or whether the capacity of one brain region is unmasked by a difference in another. High-resolution neuroimaging methods may begin to answer these questions. Functional imaging studies show that the brain's grey and white matter are both capable of considerable plasticity. Description of underlying brain mechanisms is not the province of *The Genius Within*; these will eventually deserve an authoritative account.

Adam also ventures into commercial brain games and apps, which challenge attention and memory. Computerized training of working memory (through, for instance, remembering the locations of objects in grids) has been claimed to improve cognition in both healthy people and those with mental illnesses. Given that working memory is closely related to IQ, such training might be expected to produce more-general improvements in intelligence. However, controversial attempts at controlled trials have sometimes cast doubt on this.

I enjoyed *The Genius Within* enormously. Eminently readable, it takes a relatively holistic 'mind-body' view of our abilities, beyond our capacity to decipher rotating polygons or shuffling anagrams. Dancers and sportspeople have obvious 'motor intelligence': executing complex choreography or turning a football game on a visionary pass demand visual and spatial acuity, exquisite coordination and seamless interaction with others. So Adam's discussion of how electrical stimulation of the motor cortex improves the performance of road cyclists is fascinating (even as it raises questions about detection).

IQ testing will continue to be criticized. We would do better to take into account a broader range of cognitive capacities, with more-sophisticated measures of behaviour in a range of situations. And we must work harder to obtain an optimal balance of these capacities by fine-tuning our brains through ever-evolving methods. ■

**Trevor Robbins** is professor of cognitive neuroscience at the University of Cambridge, UK.  
e-mail: twr2@cam.ac.uk



**The Genius Within: Unlocking Your Brain's Potential**  
DAVID ADAM  
Pegasus: 2018.

# Correspondence

## Existing rules cover gene-drive usage

Gene-drive technology is not unregulated, as you imply (*Nature* 552, 6; 2017). Because it involves genetically modified (GM) organisms, it is covered in countries that have regulations on gene modification and internationally by the Cartagena Protocol on Biosafety.

It could be argued that the risks are not comparable for contained laboratory use versus deliberate release of GM organisms into the wild. This assumes that lab safety standards based on pathogenicity would be inadequate for non-pathogenic gene-drive organisms. However, European regulations, as well as, for example, German law, put protection of the environment on a par with protecting human health, even for contained usage. The potential of GM organisms to persist in the environment and spread into wild populations has always been a crucial part of risk assessment for transgenic organisms.

Existing regulations therefore cover environmental risks arising from contained handling of gene-drive organisms, as confirmed by the German Central Commission for Biological Safety (see [go.nature.com/2enrjy4](http://go.nature.com/2enrjy4)). Researchers in Germany and the Netherlands need permission for gene-drive experiments. Risk assessment is then made on a case-by-case basis.

**Swantje Strassheim, Werner Schenkel** *Federal Office of Consumer Protection and Food Safety, Berlin, Germany.*  
[swantje.strassheim@bvl.bund.de](mailto:swantje.strassheim@bvl.bund.de)

## Bitcoin's alarming carbon footprint

The 'mining' process for the cryptocurrency bitcoin is power hungry, and is increasing its environmental impact as its price and popularity rise. Cryptocurrencies are generated by specialized software, used

to solve complex mathematical problems that represent proof-of-work algorithms in exchange for electronic coins (see <https://bitcoin.org/bitcoin.pdf>).

Some estimate that the combined electricity consumption for bitcoin and ethereum mining, which together represent 88% of the total cryptocurrency market capitalization (G. Hileman and M. Rauchs <http://doi.org/cj22>; 2017), has already reached a staggering 47 terawatt-hours per year and is on the rise (see [www.digiconomist.net](http://www.digiconomist.net)). To put this into perspective, Greece's population of 11 million consumes close to 57 terawatt-hours annually.

Moreover, 58% of all cryptocurrency mining is done in China and is typically powered by coal plants. Using the life-cycle impact-assessment methodology, I estimate that the annual carbon footprint for bitcoin and ethereum mining is comparable to that of some 6.8 million average European inhabitants — or as much as 43.9 million tonnes of carbon dioxide equivalent (see ReCiPe and IPCC 2013 methods, respectively, at [go.nature.com/2nn7zzj](http://go.nature.com/2nn7zzj)).

In my opinion, the cryptocurrency industry is urgently in need of reform to make it environmentally sustainable.

**Spyros Foteinis** *PPC Renewables, Athens, Greece.*  
[sfoteinis@ppcr.gr](mailto:sfoteinis@ppcr.gr)

## Baleen whale species at risk of extinction

The latest Critically Endangered list from the International Union for Conservation of Nature includes the Gulf of Mexico whale, a subspecies of *Balaenoptera edeni* (see [go.nature.com/2bdntor](http://go.nature.com/2bdntor)). This mammal is at risk of being the first baleen whale to go extinct since the Atlantic grey whale (*Eschrichtius robustus*) three centuries ago. Yet the animal's new status has generated little public response.

The Gulf of Mexico whale is similar to Bryde's and Eden's whales (both also named *B. edeni*), but is genetically distinct from both. It is entirely confined to US waters in the Gulf of Mexico (see [go.nature.com/2bdntor](http://go.nature.com/2bdntor)). Survey data put its abundance at 33 individuals in 2009, and modelling suggests that almost half its habitat was affected by the Deepwater Horizon oil spill in 2010 (see [go.nature.com/2e6joqe](http://go.nature.com/2e6joqe)).

Rapid action is needed to eliminate sources of human-induced death and injury among these whales. A first step must be to raise society's and scientists' awareness of their status.

**Peter Corkeron** *Northeast Fisheries Science Center, Woods Hole, Massachusetts, USA.*  
**Scott D. Kraus** *New England Aquarium, Boston, Massachusetts, USA.*  
[peter.corkeron@noaa.gov](mailto:peter.corkeron@noaa.gov)

## Value and reward regional research

Incentives for publishing in international journals could be preventing ecologists in low-income countries from conducting the research needed to protect and restore their local environments. Few scientists are willing to do time-consuming taxonomic surveys, for example, because these will not generate highly cited publications. Yet effective management is impossible without such local ecological insight.

Although reward structures for research vary substantially between and within countries, they are often based on scientists' publication and citation counts in internationally recognized journals. In Mexico, for instance, this encourages research that appeals to reviewers and editors in distant countries, fosters publication in journals that are financially and linguistically inaccessible, and may not be relevant to local problems (M. W. Neff *Sci. Public Policy*

<http://doi.org/cjz2>; 2017). Journals that are regionally relevant and in languages other than English suffer because top scientists eschew them, leaving university students, resource managers and policymakers with fewer resources.

Mexico's national research policies provide clear examples of distorting incentives, but the problem is close to universal: what is countable is not always what we should be counting. Scientists and publishers need to exert their power to change these systems.

**Mark Neff** *Western Washington University, Bellingham, USA.*  
[mark.neff@wwu.edu](mailto:mark.neff@wwu.edu)

## Don't belittle junior researchers

The most interesting part of a scientific seminar, colloquium or conference for me is the question and answer session. However, I find it upsetting to witness the unnecessarily hard time that is increasingly given to junior presenters at such meetings. As inquisitive scientists, we do not have the right to undermine or denigrate the efforts of fellow researchers — even when their reply is unconvincing.

It is our responsibility to nurture upcoming researchers. Firing at a speaker from the front row is unlikely to enhance discussions. In my experience, it is more productive to offer positive queries and suggestions, and save negative feedback for more-private settings.

With belligerence supplanting courtesy inside and outside the conference room, it might be helpful for young researchers to be taught how to frame a question in a purely scientific way. Let us create a system in which junior scientists feel excited to share their data.

**Anand Kumar Sharma** *CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India.*  
[anandkumar@ccmb.res.in](mailto:anandkumar@ccmb.res.in)

# Contesting early archaeology in California

ARISING FROM S. R. Holen *et al. Nature* **544**, 479–483 (2017); doi:10.1038/nature22065

The peopling of the Americas is a topic of ongoing scientific interest and rigorous debate<sup>1,2</sup>. Holen *et al.*<sup>3</sup> add to these discussions with their recent report of a 130,000-year-old archaeological site in southern California, USA: the Cerutti Mastodon (CM) site, which includes the fragmentary remains of a single mastodon (*Mammuth americanum*), spatially associated stone cobbles, and associated lithic debris that they claim indicates prehistoric hominin activity. In sharp contrast, we contend that the CM record is more parsimoniously explained as the result of common geological and taphonomic processes, and does not indicate prehistoric hominin involvement. Whereas further investigations may yet identify unambiguous evidence of hominins in California around 130,000 years ago, we urge caution in interpreting the current record. There is a Reply to this Comment by Holen, S. R. *et al. Nature* **554**, <http://dx.doi.org/10.1038/nature25166> (2018).

Holen *et al.*<sup>3</sup> claim prehistoric hominin involvement at the CM site based primarily on four lines of evidence: a reliable radiometric age; the presence of stone artefacts; clear evidence of tool-imparted percussion damage to the remains of the mastodon; and an undisturbed geological context. We take no issue with the published age for the site, but we believe that the latter three claims warrant further examination.

The CM site stone artefacts are an assortment of cobbles and fractured cobble fragments excavated from a sandy silt matrix. There is no evidence of formal stone tool forms or diagnostic lithic micro- or macro-debitage. Instead, the CM site artefacts are identified by their proximity to the remains of the mastodon and their large size relative to the enveloping sediment. Additionally, surface features such as the presence of pitting and scratching on cobble surfaces, the presence of several cobble fragments with fracture morphologies reminiscent of hammerstone and/or anvil usage, and the presence of several refitting cobble fragments are interpreted as evidence of hominin percussive activities on-site. The lack of discarded formal tools and diagnostic lithic debitage is noteworthy, and is unusual relative to most archaeological assemblages that purport hominin processing of proboscidean remains (although see Haynes<sup>4</sup>). We also note that upslope of the site there are numerous alluvial fans, with clastic material a common occurrence. The cobbles and pebbles at the CM site can be derived from modest alluvial fan input with fines subsequently winnowed with lower energy fluvial erosion. Crucially, none of the criteria that Holen *et al.*<sup>3</sup> use to define stone artefacts either requires prehistoric hominin involvement or meets the accepted criteria for falsifying natural ‘geofacts’<sup>5</sup>. The range of possible geological interpretations for the lithic assemblage highlights the critical issue of equifinality, in which an end product such as a shattered cobble may be generated via many and potentially unrelated means. It is a principle that applies to the bone record as well.

We contend that Holen *et al.*<sup>3</sup> presented equivocal evidence in support of tool-imparted percussion damage to the remains of the mastodon. The critical observations are of spiral fractures, cone flakes, impact flakes, bulbs of percussion, impact notches, negative flake scars, anvil-polished specimens, percussion-fractured specimens, and refitting specimens. These bone damage features are inferred to implicate sole agency by prehistoric hominins. As with the stone artefact record, issues of equifinality must first be addressed, including the question of whether other processes could produce such bone damage.

Haynes<sup>6</sup> presents compelling evidence of non-cultural and/or non-prehistoric processes producing comparable damage at the

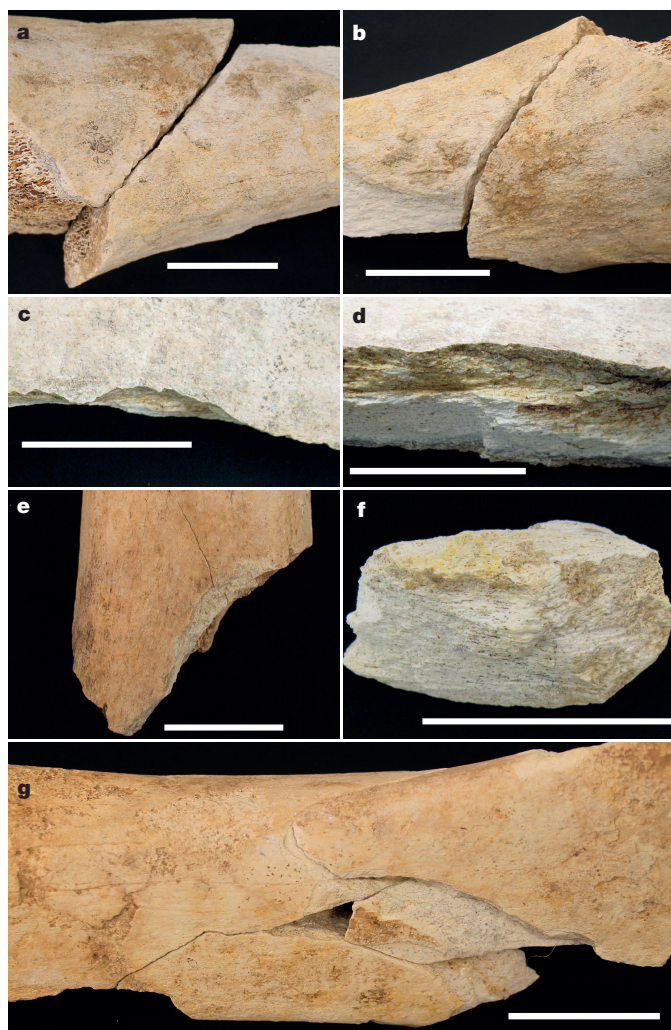
24,000-year-old Inglewood Mammoth Site (IMS), Maryland, USA. As at the CM site, the IMS contains the remains of a single juvenile proboscidean recovered *in situ* from a sealed deposit of sandy clays with pebbles and cobbles<sup>6</sup>. Haynes<sup>6</sup> provides descriptions and images of curvilinear and spiral ‘green-bone’ fractures on cranial, axial and appendicular specimens. Some of these fractures are recent in origin, probably related to heavy earthmoving equipment working on-site<sup>6</sup>. Other damage may reflect perimortem injuries sustained by the mammoth. No evidence of prehistoric hominin activities are noted or suspected for the site. Post-burial bone notches, impact points and impact scratches occurred on a number of specimens.

We report a similar record of fractured proboscidean bones at the Waco Mammoth National Monument (WMNM), Waco, Texas, USA. The site contains the remains of at least 26 mammoths buried in fluvial sands, silts and clays, and dates from 66,800 to 51,300 years ago<sup>7</sup>. The WMNM was initially investigated as a potential archaeological site, although no evidence of prehistoric human activities was discovered. Figure 1 shows post-burial damage to WMNM mammoth long bones morphologically consistent with observations from the IMS and CM sites. This includes damage that resembles spiral fractures with associated sedimentary abrasion, hammerstone pseudo-notches<sup>8</sup>, negative flake scars, partially detached flakes and incipient notches, and bulbs of percussion. Such damage, including spiral fractures, is well known in the fossil record from as early as the Triassic period<sup>9</sup> and can occur post-burial<sup>6</sup>. They neither require nor solely indicate prehistoric human agency<sup>4,6,8</sup>.

Other proboscidean assemblages share a similar taphonomic signature with the WMNM, IMS and CM sites. Holen and others report various combinations of spiral fractures, impact points, bulbs of percussion and bone flakes at numerous other late Pleistocene mammoth death sites in the Americas<sup>10,11</sup>. As with the CM site, these latter assemblages uniformly lack unambiguous stone tools, cut marks, or any other unquestionable evidence of hominin activities, and most predate well-vetted geochronological and palaeogenomic evidence of the initial peopling of the Americas around 15.5 thousand years ago<sup>1,2,12–14</sup>.

Moreover, it is not just what is present at the CM site, but also what is missing. Specifically, hammerstone striae and/or pits (HSSP)<sup>15</sup> are noticeably absent despite reasonable cortical bone preservation, several hundred bone fragments, purported hammerstones, and purported anvil abrasions on both the bones and the cobbles. Experimental studies show that hammerstone-percussed proboscidean limb bone fragments should bear HSSP on greater than 30% of specimens created when using a hafted hammerstone and anvil<sup>12</sup>. The absence of HSSP at the CM site—a proposed percussed bone assemblage—cannot be explained using current experimental models and contradicts the assumption of hammerstone-wielding hominin involvement in bone breakage.

Lastly, we question the assertion of an “undisturbed geologic context” at the CM site. Although the distance between some refitted finds necessarily suggests pre-burial breakage and scattering of some items (for example the molar fragments), other features of the record plausibly reflect subsequent forces modifying the assemblage over the last 130,000 years. As fluvial deposits slowly covered the remains, the bones of the mastodon would have remained semi-pliable for years<sup>6</sup>. Proboscideans or other large mammals subsequently using the muddy watercourse could potentially trample, displace, fracture, abrade and reorient (for example the semi-vertical tusk) the interred materials<sup>4,6</sup>. Later sediment compaction by metres of overburden and then eventual



**Figure 1 | Fractured proboscidean bones from WMNM.** **a**, Curvilinear fracture between refitted fragments BU-MMC-641a and BU-MMC-641b. **b**, The opposite side of the refitted fragments depicted in **a**. **c**, Hammerstone-like 'micro'-notch<sup>8</sup> on the cortical surface without diagnostic percussion pit, fragment BU-MMC-1011a. **d**, Negative flake scar on the same bone as that depicted in **c**. **e**, Post-burial curvilinear fracture on fragment BU-MMC-210a. **f**, Bone flake with bulb of percussion, fragment BU-MMC-642b. **g**, Comminuted fracture with refitted flakes and associated sedimentary abrasion on fragment BU-MMC-716a. Scale bars, 5 cm.

disturbance by heavy construction equipment (see supplementary information 6 of ref. 3) would confound the taphonomic history of the site further, as it has at both IMS and WMNM<sup>6</sup>.

The extraordinary claim by Holen *et al.*<sup>3</sup> of prehistoric hominin involvement at the CM site should not be contingent on evidence that is open to multiple, contrasting interpretations. Until unambiguous evidence of hominin activities can be presented, such as formal stone tools or an abundance of percussion pits, caution requires us to set aside the claims of Holen *et al.*<sup>3</sup> of prehistoric hominin activities at the CM site.

## Methods

The Baylor University Mayborn Museum Complex (BU-MMC), Waco, Texas, is the official repository for around 70% of the recovered WMNM remains, with the remainder left *in situ* for display at the WMNM site. Approximately 1,100 trays of curated fossils are available for study, with most trays containing multiple specimens. Individual specimens are labelled here based on tray number (BU-MMC), followed by a letter designation (for example 642a, 642b). Specimens were selected based on gross bone damage morphologies, with the aim of recording damage similar to that reported from the CM site. Images were obtained using a Cannon EOS Rebel XS digital camera.

**Data availability.** All data are available from the corresponding author upon reasonable request.

**Joseph V. Ferraro<sup>1</sup>, Katie M. Binetti<sup>1</sup>, Logan A. Wiest<sup>2</sup>, Donald Esker<sup>2</sup>, Lori E. Baker<sup>1</sup> & Steven L. Forman<sup>2</sup>**

<sup>1</sup>Department of Anthropology and Institute of Archaeology, Baylor University, Waco, Texas, USA.

email: joseph\_ferraro@baylor.edu

<sup>2</sup>Terrestrial Paleoclimatology Research Group, Department of Geosciences, and Institute of Archaeology, Baylor University, Waco, Texas, USA.

**Received 7 July; accepted 8 November 2017.**

1. Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
2. Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**, 1497–1502 (2008).
3. Holen, S. R. *et al.* A 130,000-year-old archaeological site in southern California, USA. *Nature* **544**, 479–483 (2017).
4. Haynes, G. *Mammoths, Mastodons, and Elephants: Biology, Behavior, and the Fossil Record* (Cambridge Univ. Press, 1991).
5. Mercader, J. *et al.* 4,300-year-old chimpanzee sites and the origins of percussive stone technology. *Proc. Natl Acad. Sci. USA* **104**, 3043–3048 (2007).
6. Haynes, G. Taphonomy of the Inglewood mammoth (*Mammuthus columbi*) (Maryland, USA): green-bone fracturing of fossil bones. *Quat. Int.* **445**, 171–183 (2017).
7. Wiest, L. A., Esker, D. & Driese, S. G. The Waco Mammoth National Monument may represent a diminished watering-hole scenario based on preliminary evidence of post-mortem scavenging. *Palaio* **31**, 592–606 (2016).
8. Capaldo, S. D. & Blumenschine, R. J. A quantitative diagnosis of notches made by hammerstone percussion and carnivore gnawing on bovid long bones. *Am. Antiq.* **59**, 724–748 (1994).
9. Hunt, A. P. & Lucas, S. G. in *Fossil Vertebrates in New Mexico* (eds Lucas, S. G. & Sullivan R. M.) 9–40 (New Mexico Museum of Natural History and Science Bulletin 68, 2015).
10. Holen, S. R. Taphonomy of two last glacial maximum mammoth sites in the central Great Plains of North America: a preliminary report on La Sena and Lovewell. *Quat. Int.* **142–143**, 30–43 (2006).
11. Holen, S. R. & Holen, K. in *IV Simposio Internacional El Hombre Temprano en América* (eds Jimenez Lopez, J. C. *et al.*) 85–105 (Instituto Nacional de Antropología e Historia, 2012).
12. Krasinski, K. E. *Broken Bones and Cutmarks: Taphonomic Analyses and Implications for the Peopling of North America*. PhD thesis, Univ. of Nevada, Reno (2010).
13. Waters, M. R. *et al.* The Buttermilk Creek complex and the origins of Clovis at the Debra L. Friedkin site, Texas. *Science* **331**, 1599–1603 (2011).
14. Grayson, D. K. & Meltzer, D. J. Revisiting Paleoindian exploitation of extinct North American mammals. *J. Archaeol. Sci.* **56**, 177–193 (2015).
15. Blumenschine, R. J. & Selvaggio, M. M. Percussion marks on bone surfaces as a new diagnostic of hominid behaviour. *Nature* **333**, 763–765 (1988).

**Author Contributions** J.V.F., K.M.B. and S.L.F. conceived the project; J.V.F., K.M.B., L.A.W., D.E. and L.E.B. interpreted the taphonomic record; S.L.F. interpreted the geological record; J.V.F. and S.L.F. interpreted the lithic technology record; J.V.F. photographed the WMNM mammoth bones; L.A.W. created the figures; J.V.F., K.M.B. and S.L.F. wrote the manuscript; J.V.F., K.M.B., S.L.F., L.A.W., D.E. and L.E.B. edited the manuscript.

**Competing Financial Interests** Declared none.

doi:10.1038/nature25165

Holen *et al.* replyREPLYING TO J. V. Ferraro *et al.* *Nature* **554**, <http://dx.doi.org/10.1038/nature25165> (2018)

Contrary to our hypothesis<sup>1</sup> that the Cerutti Mastodon (CM) site represents a 130,000-year-old archaeological site, in the accompanying Comment<sup>2</sup> Ferraro *et al.* argue that the site formed through ‘common’ geological and taphonomic processes. As a source for the cobbles that we interpreted as hammerstones and anvils, they postulate a previously unrecognized alluvial fan, later removed by fluvial winnowing that somehow left our five cobbles, refitting flakes, and fragments of stone, mastodon bone and teeth in place. There is no sedimentological or geomorphic evidence of an alluvial fan, and their scenario leaves unexplained a number of taphonomic features, including the two discrete concentrations in which were found cobbles, refit stones and bones, impact-fractured bones, side-by-side femoral heads and a tusk oriented vertically.

Ferraro *et al.*<sup>2</sup> also speculate that the stone and bone fractures that we analysed can be explained by post-burial processes such as sediment compaction or interaction with excavation equipment, whereas we contend that these features are part of the CM biostratigraphic (pre-burial) record. Support for our view is provided by the fact that most CM bones and stones were enclosed within crusts of pedogenic carbonate that establish a ‘chain of evidence’ showing that breakage and positioning of objects occurred many thousands of years ago, and, as we contend, before burial<sup>3</sup>. The only pre-burial cause of bone breakage Ferraro *et al.*<sup>2</sup> consider is trampling, which we have argued is incompatible with other site data<sup>1</sup>.

Ferraro *et al.*<sup>2</sup> draw comparisons to the Inglewood Mammoth Site (IMS)<sup>4,5</sup> and the Waco Mammoth National Monument (WMNM)<sup>6</sup>. For the IMS, they cite an observationally based study<sup>4</sup> that proposes that excavating equipment caused the spiral fractures on many of the bones. However, this claim is compellingly refuted by an experimentally based study<sup>5</sup> that shows that the IMS spiral fractures are ancient after all, and probably occurred before burial.

WMNM bones illustrated by Ferraro *et al.* (figure 1 of ref. 2) lack clear evidence of true spiral fractures or normal impact notches<sup>7</sup>, instead representing classic examples of dry bone fracture, with rough texture on fracture surfaces and contrasting coloration of broken versus cortical surfaces (figure 1b, d, g of ref. 2). The closest approach to a notch (shown in figure 1c, d of Ferraro *et al.*<sup>2</sup>) is a shallow, irregularly arcuate break—described as a pseudo-notch or micro-notch—that does not extend to the medullary portion of the bone, unlike the ‘normal notch’<sup>7,8</sup> we illustrated<sup>1</sup>, which was defined by two clear inflection points, a negative flake scar, an attached cone flake and smoothly curved fracture surfaces that extend completely through the cortical portion of the bone. Only ‘normal notches’ are used to determine human agency<sup>7,8</sup>.

By overlooking the most important bone evidence, which includes impact features such as cone flakes, bulbs of percussion and a large impact notch with associated negative flake scar, as well as bone distribution patterns, bone refits and missing femoral diaphysis pieces,

Ferraro *et al.*<sup>2</sup> did not consider precisely those features that are individually and collectively most likely to have been caused by cultural processes. They have not offered a cogent alternative site formation hypothesis that accounts for all evidence presented.

**Steven R. Holen<sup>1,2</sup>, Thomas A. Deméré<sup>1</sup>, Daniel C. Fisher<sup>3,4</sup>, Richard Fullagar<sup>5</sup>, James B. Paces<sup>6</sup>, George T. Jefferson<sup>7</sup>, Jared M. Beeton<sup>8</sup>, Richard A. Cerutti<sup>1</sup>, Adam N. Rountrey<sup>3</sup>, Lawrence Vescera<sup>7</sup> & Kathleen A. Holen<sup>1,2</sup>**

<sup>1</sup>Department of Paleontology, San Diego Natural History Museum, San Diego, California, USA.

email: [tdemere@sdnhm.org](mailto:tdemere@sdnhm.org)

<sup>2</sup>Center for American Paleolithic Research, 27930 Cascade Road, Hot Springs, South Dakota, USA.

<sup>3</sup>Museum of Paleontology, University of Michigan, Ann Arbor, Michigan, USA.

<sup>4</sup>Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA.

<sup>5</sup>Centre for Archaeological Science, School of Earth and Environmental Sciences, Faculty of Science Medicine and Health, University of Wollongong, Wollongong, New South Wales, Australia.

<sup>6</sup>Geosciences and Environmental Change Science Center, US Geological Survey, Denver, Colorado, USA.

<sup>7</sup>Colorado Desert District, Stout Research Center, California Department of Parks and Recreation, Borrego Springs, California, USA.

<sup>8</sup>Department of Earth Science, Adams State University, Alamosa, Colorado, USA.

- Holen, S. R. *et al.* A 130,000-year-old archaeological site in southern California, USA. *Nature* **544**, 479–483 (2017).
- Ferraro, J. V. *et al.* Contesting early archaeology in California. *Nature* **554**, <http://dx.doi.org/10.1038/nature25165> (2018).
- Holen, S. R. *et al.* Broken bones and hammerstones at the Cerutti Mastodon site: a reply to Haynes. *PaleoAmerica* <https://doi.org/10.1080/20555563.2017.1396835> (2017).
- Haynes, G. Taphonomy of the Inglewood Mammoth (*Mammuthus columbi*) (Maryland, USA): green-bone fracturing of fossil bones. *Quat. Int.* **445**, 171–183 (2017).
- Karr, L. P. Human use and reuse of megafaunal bones in North America: bone fracture, taphonomy, and archaeological interpretation. *Quat. Int.* **361**, 332–341 (2015).
- Wiest, L. A., Esker, D. & Driese, S. G. The Waco Mammoth National Monument may represent a diminished watering-hole scenario based on preliminary evidence of post-mortem scavenging. *Palaeos* **31**, 592–606 (2016).
- Capaldo, S. D. & Blumenschine, R. J. A quantitative diagnosis of notches made by hammerstone percussion and carnivore gnawing on bovid long bones. *Am. Antiq.* **59**, 724–748 (1994).
- Pickering, T. R. & Egeland, C. P. Experimental patterns of hammerstone percussion damage on bones: Implications for inferences of carcass processing by humans. *J. Archaeol. Sci.* **33**, 459–469 (2006).

doi:10.1038/nature25166

## MATERIALS SCIENCE

## Wood made denser and stronger

An improved method for compressing wood substantially increases its strength and stiffness, opening up the possibility of applications in engineering for which natural wood is too weak. [SEE LETTER P.224](#)

PETER FRATZL

Wood is among the oldest materials used by humans, and is still commonly used for building<sup>1</sup>. Its low density has also made it useful for transport applications such as shipbuilding, but this property is associated with a relatively low strength and stiffness. Scientists have tried to devise processes that make wood denser, to obtain materials suitable for high-strength applications, but with limited success. On page 224, Song *et al.*<sup>2</sup> describe a densification method that combines a chemical treatment with high-temperature compression, and which produces an unprecedented increase in stiffness and strength.

The authors' method starts by treating wood blocks with sodium hydroxide and sodium sulfite, a chemical process similar to the method used to pulp wood to make paper. This chemical treatment partially removes lignin and hemicelluloses (Fig. 1). Lignin is a biopolymer that has many functions in plants, such as stabilizing cell walls in wood and retarding attacks on wood cells by parasites and bacteria; hemicelluloses are sugar chains that cover and bind fibrils of cellulose in the cell walls.

Song and colleagues then compress the blocks at temperatures of about 100 °C. This removes most of the pores in the wood, and

increases its density from 0.43 grams per cubic centimetre to 1.3 g cm<sup>-3</sup>. The resulting stable material is too dense to float on water, but the authors report that its stiffness and strength have both increased impressively, by a factor of about 11 compared with untreated wood. As the authors point out, previous attempts to densify wood also improved the strength, but by no more than a factor of about three to four<sup>3</sup>. The secret to Song and colleagues' success lies in their combination of chemical treatment and high temperatures during pressing.

Natural wood contains a multitude of parallel, tube-like cells, the walls of which constitute the major part of the material. In most parts of woody stems, the cells have died and left behind their cellulose-rich cell walls. These walls also contain lignin and hemicelluloses, and form hollow wood fibres. The tube-like fibres collapse laterally when compacted, effectively losing their hollow interiors. This increases the amount of material per cross-section of the stem, as evidenced by the increased density reported by Song and colleagues. On its own, this effect would be expected to cause the stiffness and strength of wood to increase in proportion to the increase in density<sup>4</sup>.

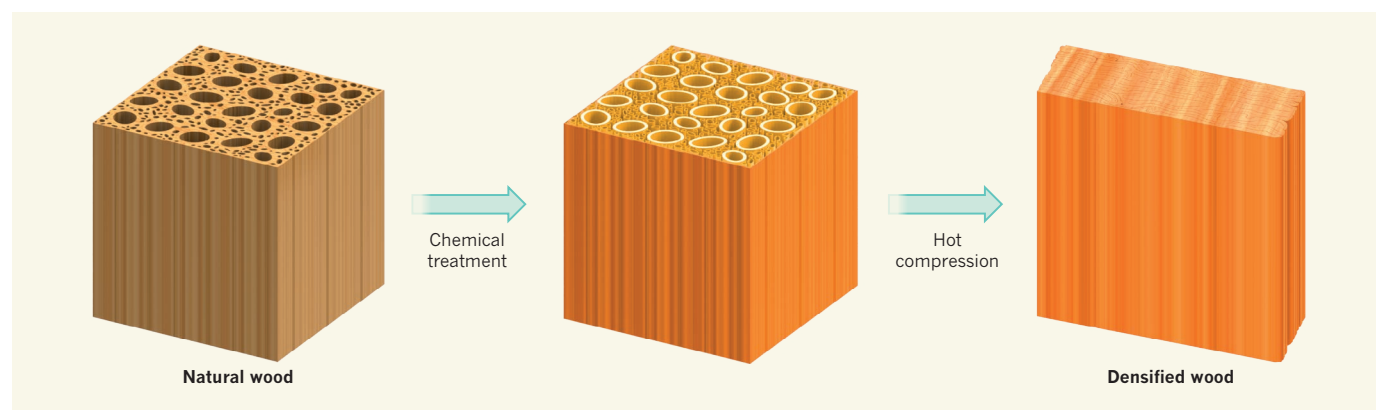
However, the authors report that the stiffness increases by a factor of 11, whereas the density increases by a factor of only 3. A threefold density increase has been observed

in previous work that used hot pressing alone (see ref. 3, for example). It therefore seems likely that the authors' chemical treatment modifies and strengthens the cellulose-based composite that makes up cell walls in wood.

Many cellulose-based materials swell undesirably when they come into contact with water, but Song and colleagues report that water swelling of their densified wood is tolerably small. It remains to be seen whether the partial removal of lignin from the material makes it susceptible to bacterial or fungal attacks.

The densified wood is still lighter than metallic materials, so its stiffness and strength open up the potential for many engineering applications. This raises the question of why trees use a porous material for their trunks, when their goal in a forest is to be as high as possible, to ensure that their leaves are exposed to light — a task for which stiffer and stronger materials might intuitively seem better suited. By making wood porous, trees partially sacrifice the material's strength. One answer is that wood is multifunctional, and so the pores are needed for more than just structural tasks, such as to transport water and nutrients.

But the optimal response of natural materials to a load varies according to the function involved, such that lower density can be more important than higher strength<sup>5-7</sup>. In brief, the height of a slender column that supports a



**Figure 1 | A process for densifying wood.** Natural wood contains pores formed from the remains of parallel, tube-like cells, the walls of which contain cellulose, along with biopolymers known as lignin and hemicelluloses. Song *et al.*<sup>2</sup> treated natural wood with a mixture of sodium hydroxide and sodium sulfite, which partly removed the lignin

and hemicelluloses. They then compressed the wood at about 100 °C, which caused the cells to collapse. The resulting material was about 3 times as dense as natural wood, and about 11 times as stiff and strong — making it potentially useful for high-strength engineering applications. (Adapted from ref. 2.)

compressive load along its axis is often limited by the risk of buckling; for a given column width and compressive load, higher columns can be built by using materials that have a higher Young's modulus ( $E$ , a measure of stiffness). When the height of the column is not limited by an external load, but by just its own weight, then greater heights can be attained using a less dense material: the aim in this context is to maximize the ratio of  $E$  to the density  $\rho$ , rather than just  $E$ . And when the goal is to build the highest possible column using a fixed mass of material, then it is best to maximize  $E/\rho^2$ . (Maintaining a constant mass is relevant to plants, because synthesizing material is a major cost for them; maximizing  $E/\rho^2$  corresponds to the most economical way of growing the highest possible column at fixed material costs.)

A consideration of these principles reveals that Song and colleagues' densified wood should perform better than natural, porous wood in the first two scenarios (in which  $E$  or  $E/\rho$  need to be as large as possible), but only about equally well in the third situation, for which  $E/\rho^2$  is maximized, on the basis of the changes in stiffness and density reported by the authors. This indicates that trees do not lose much by making wood porous, and that the introduction of pores for water transport comes at no extra material cost. Perhaps because of this, the height of trees is likely to be limited more by hydraulic constraints linked to water transport than by mechanical constraints<sup>8</sup>. Similarly, many advanced-engineering applications require materials that have high stiffness and strength, but in some cases porous materials would increase performance, rather than decrease it.

All biological materials are active, and adapt their internal structure to their function and to environmental needs. Two strategies can be used to repurpose such materials for engineering applications. One is to modify the material to comply with specifications in industrial design, as exemplified by Song *et al.* with their densification procedure. The other, perhaps more conventional, option is to adapt designs to the properties of natural materials. The latter approach is more sustainable, but would require greater knowledge of how structure relates to function in such materials, and the development of new design approaches<sup>9,10</sup>. ■

**Peter Fratzl** is in the Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Potsdam 14424, Germany. e-mail: fratzl@mpikg.mpg.de

1. Ramage, M. H. *et al. Sustain. Energ. Rev.* **68**, 333–359 (2017).
2. Song, J. *et al. Nature* **554**, 224–228 (2018).
3. Erickson, E. *Mechanical Properties of Laminated Modified Wood* (US Dept Agriculture, 1965).
4. Ashby, M. F. *Metall. Trans. A* **14**, 1755–1769 (1983).
5. Ashby, M. F., Gibson, L. J., Wegst, U. & Olive, R. *Proc. R. Soc. Lond. A* **450**, 123–140 (1995).
6. Gibson, L. J., Ashby, M. F., Karam, G. N., Wegst, U. & Shercliff, H. R. *Proc. R. Soc. Lond. A* **450**, 141–162 (1995).
7. Fratzl, P. & Weinkamer, R. *Prog. Mater. Sci.* **52**,

- 1263–1334 (2007).
8. Niklas, K. J. *Tree Physiol.* **27**, 433–440 (2006).
9. Schäffner, W. in *Ultra Knowledge & Gestaltung* (eds Doll, N., Bredekamp, H. & Schäffner, W.) 23–32

- (Seemann, 2017).
10. Fratzl, P. in *Ultra Knowledge & Gestaltung* (eds Doll, N., Bredekamp, H. & Schäffner, W.) 173–178 (Seemann, 2017).

## CANCER RESEARCH

# Many mutations in one clinical-trial basket

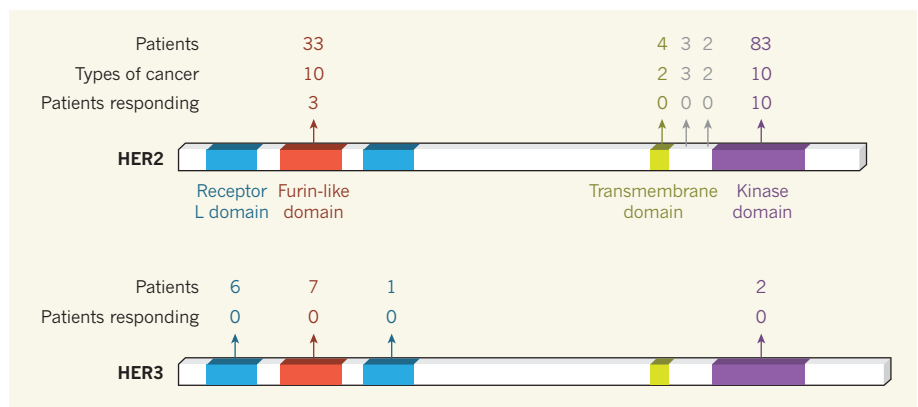
**When abnormality in a gene is linked to cancer and a drug targets the encoded protein, how can the patients who will respond to the drug be identified if the gene is mutated in many different ways in many different cancers?** [SEE ARTICLE P.189](#)

ELAINE R. MARDIS

Cancer usually arises from genomic abnormalities. However, the number and complexity of genetic alterations in tumours can make it difficult to predict whether, and in which tissues, a particular mutation in a specific cancer-linked gene will drive tumour growth. This poses a challenge when trying to identify effective treatments. For example, if a drug that targets a specific protein can treat a person with breast cancer who has a mutation in the gene encoding the protein, could the drug treat another patient who has a different mutation in that gene? And could it treat a person with a mutation in the same gene, but in a tumour that has developed in a different tissue? On page 189, Hyman *et al.*<sup>1</sup> report the outcome of a clinical trial testing the ability of the drug neratinib, which inhibits HER2 and HER3 tyrosine kinase enzymes, to reduce or eliminate tumours. The drug was

tested on 21 types of cancer in 141 people who had a total of 42 different mutations affecting one of the enzymes.

Studies in the 1970s revealed that certain chromosomal DNA aberrations can be linked to the development of specific cancer types, and that an amplification in the number of copies of particular genes can have a tumour-promoting effect<sup>2</sup>. For example, a highly lethal type of breast cancer is linked<sup>3</sup> to amplification of the gene *ERBB2* and an increase in the level of the HER2 protein that it encodes. HER2 amplification occurs in several other cancers<sup>4</sup>, including colorectal adenocarcinoma and bladder cancer. This understanding led to efforts to develop treatments to stop the action of such overexpressed proteins, resulting in several HER2-targeted therapies that are used in the clinic<sup>5</sup> to prolong survival in people whose cancers have amplification of *ERBB2*. Other links between *ERBB2* abnormalities and cancer have been identified;



**Figure 1 | Results of a cancer clinical trial.** Hyman *et al.*<sup>1</sup> report the outcome of a study testing how effectively the drug neratinib can treat tumours. The tyrosine kinase enzymes HER2 and HER3 have been linked to tumour growth and can be inhibited by neratinib. The 141 patients tested had a range of mutations that altered HER2 or HER3, and, between them, had many different tumour types. The protein structures are shown, and arrows indicate the domains or interdomain locations at which protein alterations due to mutations were found. For the HER2 data shown, the cancers were grouped into ten cancer-type categories: biliary, bladder, breast, cervical, colorectal, endometrial, gastro-oesophageal, lung, ovarian or other (for all other cancer types). Responding patient numbers indicate those whose best overall response to the drug was a partial or complete response — a decrease or absence, respectively, of detectable cancer at the end of the trial.

for example, single-nucleotide mutations in *ERBB2* have been found in breast cancers<sup>6</sup> that do not have amplified *ERBB2*, and in lung adenocarcinomas<sup>7</sup>.

The rapid development of therapeutics targeting specific cancer-associated proteins has coincided with the rise in DNA sequencing of tumours. In the past decade, the genomic alterations in tens of thousands of cancers have been characterized at single-nucleotide resolution. This has revealed that cancer-associated genes can be altered in myriad ways and that such alterations can be found in primary tumours that arise in many different tissues. However, such variability makes it hard to predict whether a specific drug will have an effect on a patient's cancer; this, in turn, complicates the decision of who to enrol in a clinical trial. One approach to this problem involves introducing the mutated genes in question into preclinical model systems such as genetically engineered mice or cell-line models, but these models are not practical for large-scale investigations of many different gene alterations in different tissue types.

The design of clinical trials testing targeted therapeutics has changed substantially in the era of cancer genomics. Early-phase trials, in particular, now often include people who have an altered target gene, regardless of the tissue in which the tumour is present. These 'basket' trials seek to identify the combination of mutations and tissues that respond to treatment, offering the opportunity, if a trial progresses to a later stage, to focus on tumours in those tissues that are most likely to respond.

The ability of neratinib to target tumours with *ERBB2* mutations had been demonstrated<sup>6</sup> in human-tumour samples transplanted into mice. Hyman *et al.* used a basket-trial approach to test the effects of the drug on many patients with known tumour-driving *ERBB2* mutations; they also examined its effects on a small number of patients who had either rare *ERBB2* mutations or mutations in *ERBB3*, the gene that encodes HER3 and that has also been linked to tumour growth<sup>8</sup>. An interesting feature of the trial design is that it included people with mutations that had not previously been tested for a response to the drug. Some tumour types studied by Hyman and colleagues were not represented in sufficient numbers for the team to assess whether treatment had had a statistically significant effect, and enrolment in the trial is continuing for specific tissues.

The authors found that the effect of neratinib therapy varied in different mutational and tissue contexts. For example, some people who had breast, small-cell lung, cervical, biliary or salivary cancers, and who had certain *ERBB2* mutations, responded to the treatment; the greatest effect was observed for breast cancers containing amino-acid alterations in the extracellular or kinase domains of HER2 (Fig. 1). Several patients with previously

uncharacterized *ERBB2* variants responded to neratinib, supporting the role of these mutations as tumour drivers. Neratinib had no effect on tumours with *ERBB3* mutations, nor did it affect colorectal or bladder cancers that had *ERBB2* mutations. The bladder-cancer result is consistent with previous studies<sup>9,10</sup> in which HER2 targeting did not affect this type of cancer. Lack of response to neratinib provides circumstantial evidence that rare alterations in *ERBB2* are unlikely to be tumour drivers.

Hyman and colleagues' results indicate that preclinical model studies, such as those suggesting that *ERBB3* can drive tumour growth<sup>8</sup>, can sometimes be misleading when trying to infer what happens in a human tumour. This might be because of how the overall genomic context influences the effect of a mutation. A tumour that has an altered target gene could also have alterations in other cancer-promoting genes. Another source of inconsistency between human and mouse studies might be the particular tissue context.

Finally, the genomic heterogeneity of tumour cells (the presence of groups of cells in the tumour that contain different genetic alterations) might be important in determining treatment response. Sequencing analysis conducted by Hyman and colleagues for certain *ERBB2* mutations demonstrated that most patients whose *ERBB2* mutations were present in all the tumour cells responded to neratinib, whereas those with *ERBB2* mutations in only a subset of the tumour cells did not respond.

The authors noted that response to treatment could be affected by the particular genetic mutation, the location of the tumour and the specific pattern of other mutated

cancer-associated genes present. This will probably hold true for most, if not all, future basket trials of targeted inhibitor therapies and is quite instructive for such studies. More-complete genomic characterization of tumours, beyond the gene(s) being targeted, will be needed to determine the genomic context linked to response or resistance to treatment. The genomic profiles and therapeutic-response data from basket trials such as this one should be made publicly available as a way of improving the design of clinical trials of other agents. Such data sets might contribute to the development of diagnostics that enable the precise identification of those patients who are most likely to benefit from targeted treatment. The data could also help to streamline the design of clinical trials and thereby hasten cancer therapeutics towards regulatory approval. ■

**Elaine R. Mardis** is at the Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio 43205, USA.  
e-mail: elaine.mardis@nationwidechildrens.org

1. Hyman, D. M. *et al.* *Nature* **554**, 189–194 (2018).
2. Semba, K. *et al.* *Proc. Natl Acad. Sci. USA* **82**, 6497–6501 (1985).
3. Slamon, D. J. *Science* **235**, 177–182 (1987).
4. Scholl, S., Beuzeboc, P. & Pouillart P. *Ann. Oncol.* **12**, S81–S87 (2001).
5. Slamon, D. J. *et al.* *N. Engl. J. Med.* **344**, 783–792 (2001).
6. Bose, R. *et al.* *Cancer Discov.* **3**, 224–237 (2013).
7. Stephens, P. *et al.* *Nature* **431**, 525–526 (2004).
8. Jaiswal, B. S. *et al.* *Cancer Cell* **23**, 603–617 (2013).
9. Oudard, S. *et al.* *Eur. J. Cancer* **51**, 45–54 (2015).
10. Powles, T. *et al.* *J. Clin. Oncol.* **35**, 48–55 (2017).

This article was published online on 31 January 2018.

## CLIMATE SCIENCE

## Fossil-fuel subsidies assessed

**Many governments subsidize the production and consumption of fossil fuels. Contrary to expectation, a study finds that removing these subsidies would only modestly reduce global carbon dioxide emissions. [SEE LETTER P.229](#)**

IAN PARRY

**T**he 2015 Paris climate agreement was signed by 195 countries, with most pledging to reduce their emissions of carbon dioxide and other planet-warming gases. Many countries have a long history of subsidizing fossil fuels, and it seems logical that removing these subsidies — as the G20 group of nations has agreed to do — would help them to achieve their Paris climate commitments. However, on page 229, Jewell *et al.*<sup>1</sup> report

a comprehensive and convincing analysis suggesting that reforming these subsidies would cause only a modest reduction in global CO<sub>2</sub> emissions. Nevertheless, I think that there is an urgent need for broader reform of fossil-fuel prices to fully reflect the costs associated with global warming and other environmental considerations.

Subsidy reform would increase domestic fossil-fuel prices to match the production costs. Its impact on the climate would therefore depend on how energy demand is affected by

these higher fuel prices — for example, through people driving less, power generators switching to cleaner fuels such as those from renewable energy sources, and households and businesses adopting energy-saving technologies. Because these responses are inherently uncertain, Jewell *et al.* used five different models to assess the consequences of subsidy reform. These models compared projections of fuel use and CO<sub>2</sub> emissions with and without subsidy reform by region or country, using diverse assumptions about future economic growth, technological trends, energy prices and so on.

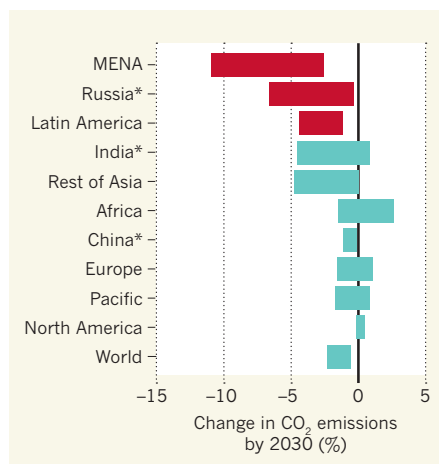
The authors found that removing all fossil-fuel subsidies would have a limited impact on global energy demand by 2030 (a reduction of about 1–4%). In addition, the share of energy from renewable sources would rise by less than 2%, and global CO<sub>2</sub> emissions would fall by only 1–4% (under both low and high oil prices). Consequently, in most regions, the CO<sub>2</sub> reduction from subsidy reform would fall far short of what is needed to meet the Paris climate pledges (Fig. 1). The exceptions are regions such as Russia, the Middle East and North Africa, where subsidies are heavily concentrated and pledges are less ambitious.

There are two main reasons for the generally modest impact of subsidy reform on CO<sub>2</sub> emissions. The first is that coal (the fossil fuel that emits by far the most CO<sub>2</sub> per unit of energy) currently receives little subsidy. Instead, 60% of subsidies are for oil, and the remainder is largely for natural gas and for the electricity generated from fuels (see Figure 2a of the paper<sup>1</sup>). The second reason is that global subsidies have declined sharply, from US\$570 billion in 2013 to \$330 billion in 2015.

However, I think that reform of fossil-fuel prices needs to go well beyond aligning them with production costs. Fuel prices should also reflect the consequences of their use for global warming and other environmental considerations, such as the costs of deaths resulting from air pollution and, in the case of road fuels, traffic congestion and accidents. Furthermore, prices for fuels purchased by households should include the general sales or value-added taxes that are applied to other consumer products.

A study<sup>2</sup> in 2017 estimated that if fossil-fuel subsidies had been defined more broadly to reflect undercharging for environmental costs and general taxes, as well as production costs, these subsidies would have totalled \$5.3 trillion in 2015 (6.5% of global gross domestic product). Furthermore, the study suggested that if prices had fully accounted for production costs, global and domestic environmental impacts and general taxes in 2013, global CO<sub>2</sub> emissions would have been 21% lower than they were, air-pollution deaths associated with fossil fuels would have been 55% lower, and government revenues as a percentage of gross domestic product would have been 4% higher.

Broader reform of fossil-fuel prices is



**Figure 1 | Impact of fossil-fuel-subsidy reform.** Fossil fuels are subsidized in many countries, and it was thought that removing these subsidies would lead to a substantial reduction in carbon dioxide emissions. However, Jewell *et al.*<sup>1</sup> report an analysis suggesting that the resulting change in CO<sub>2</sub> emissions by 2030 would be modest. The exceptions are regions in which current subsidies are heavily concentrated (shown in red), such as the Middle East and North Africa (MENA). The bars denote the range of emission changes predicted (under low oil prices), and asterisks indicate regions that constitute more than the designated country. (Adapted from Fig. 3b of ref. 1.)

therefore urgent for both developed and developing countries. However, such a reform must be carefully crafted to enhance the prospects for success. A comprehensive plan should be developed, in consultation with stakeholders, that has clear goals and timetables. It should specify the taxes to be cut or the public investment programmes to be expanded, using

revenue raised by fuel-price reform. In addition, there should be measures to compensate low-income households for the effects of higher energy prices and to help workers who might lose their jobs in energy-intensive industries.

Researchers and international organizations (such as the International Monetary Fund, World Bank and the Organisation for Economic Co-operation and Development) have an important role in providing information and guidance to help policymakers drive forward subsidy reform and communicate the case for reform to the public. The information required includes the fossil-fuel prices that countries should adopt, both to meet their Paris climate pledges and to reflect the broader environmental costs.

But it also includes the effect of reform on energy systems, the economy, fiscal balances and vulnerable groups, and the trade-offs between higher fuel prices and other policy approaches, such as requirements for energy efficiency and renewable fuels. Analysis of ongoing reform experiences in different countries could also help governments to navigate around the political obstacles.

Rigorous studies, such as that by Jewell and colleagues, are essential. But there is a need to focus these studies on the broader reform issues discussed here, for which the stakes are especially high. ■

**Ian Parry** is in the Fiscal Affairs Department, International Monetary Fund, Washington DC 20431, USA.  
e-mail: iparry@imf.org

1. Jewell, J. *et al.* *Nature* **554**, 229–232 (2018).
2. Coady, D., Parry, I., Sears, L. & Shang, B. *World Dev.* **91**, 11–27 (2017).

## BIOMECHANICS

# Evolutionary race as predators hunt prey

**Remote-sensing data for wild animals such as lions reveal that predators and prey optimize manoeuvrability rather than speed during the hunt. [SEE ARTICLE P.183](#)**

**ANDREW A. BIEWENER**

**T**he survival of predators and prey depends on their respective abilities to successfully chase food and escape capture, thereby exerting strong selective pressure on their running ability and behavioural strategies. Perhaps nowhere on Earth does this play out more dramatically than on the African savannah, where the fastest terrestrial predators chase their fleet-footed prey. Yet direct measures

of the key factors driving this type of hunt performance in the wild are difficult to obtain. On page 183, Wilson *et al.*<sup>1</sup> report findings from their use of data-capturing collars to track the movement dynamics of wild animals in Botswana during hunts. The authors also conducted computer modelling of predator–prey interactions and carried out laboratory tests to assess the properties of the animals' muscles.

In recent years, the ability to use remote-sensing devices under natural field



**Figure 1 | A lioness hunting a zebra in Etosha National Park, Namibia.** Wilson *et al.*<sup>1</sup> report their analysis of the movement dynamics of predator–prey hunts in the wild in Africa using data gathered remotely from Global Positioning System sensing collars placed on lions, zebras, cheetahs and impala.

conditions and over long time frames has led many to study animals' migratory<sup>2,3</sup>, foraging<sup>4</sup> and collective-movement behaviour<sup>5,6</sup>, which has provided fascinating insights into biomechanics, physiology and decision-making. Wilson and colleagues took a remote-sensing approach to study lions preying on zebras, and cheetahs preying on impala, in the wild. The authors temporarily immobilized animals and fitted them with lightweight collars containing technically sophisticated, custom-designed, miniature electronic and Global Positioning System (GPS) devices. The devices monitored the animals' location, movement direction and acceleration patterns. Wilson *et al.* tracked 9 lions, 5 cheetahs, 7 zebras and 7 impala, and recorded 2,726 high-speed runs for lions, 520 for cheetahs, 1,801 for zebras and 515 for impala. This remarkable data set logs individual animal strides and provides information about the speed, acceleration and turning performance of these predator–prey pairs.

The animals were not observed directly, and one limitation of the recorded data is that few, if any, of the movement tracks represented hunts between pairs of predator and prey, with both animals recorded as one hunts the other. Therefore, the hunting strategies of predator and prey must be inferred from the collar-recorded data, making the assumption that the movement patterns represent actual hunts. However, the locomotor performance

recorded by the remote-sensing collars and the hunting strategies that could be inferred from these measurements are consistent with behavioural observations made by others<sup>7</sup>. Moreover, analysis of the full data set revealed that predators and prey exhibited manoeuvrability near the limits of their capability. Hence, although recordings of one-on-one hunts are lacking, the data were consistent with maximal predator-pursuit and prey-evasion performance, enabling the authors to model hunt outcomes.

After collar placement, a tiny biopsy of hindlimb muscle was taken from the animals for subsequent state-of-the-art laboratory testing of single-muscle-fibre contractility. This revealed that, compared with the muscle fibres sampled from the prey species, the predator muscle fibres deliver more power for a given muscle mass when they contract, allowing the predators to run faster and accelerate and decelerate more quickly than their prey. With more-powerful muscles than their prey and claws to grip the ground effectively, predators are better at accelerating into a turn (centripetal acceleration) than their prey are.

Wilson and colleagues' acceleration and GPS recordings indicated that, during inferred hunts, the predators and prey regularly achieved their maximal turning performance but ran at speeds well below their athletic capabilities. Running at speeds slower than maximum capacity during a pursuit enhances

manoeuvrability, which improves the prey's probability of successful escape and enables predators to better track their prey's movements, thereby increasing the number of successful hunts.

Using their field-recorded locomotion data, Wilson and colleagues modelled predator and prey capture–evasion tactics to examine how different performance metrics, such as speed, separation distance between the animals, deceleration, acceleration and turning rate, would affect the outcome of a hunt. Evasion modelling showed that prey escape was more likely if a prey animal relied on turning more sharply and at a greater rate than its pursuer. This type of behaviour increases the unpredictability of the prey's movement trajectory, as has also been observed for bipedal desert rodents fleeing a predator<sup>8</sup>. Wilson *et al.* noted that, during the predators' approach (Fig. 1), they exhibited greater deceleration and acceleration than that of the prey, allowing the predators to close in on and better track the prey's lateral movements. The close match of athletic performance between predators and prey highlights the strong selection pressure that has resulted in an evolutionary 'arms race' for improved locomotion ability in large carnivores and their large herbivorous prey.

The increasing use of remote-sensing technologies in animal studies is enabling the monitoring of factors such as animal

acceleration, pressure (for example, during flight or when swimming at depth) and temperature. Such work promises to illuminate not only predator–prey interactions, but also how wild animals cope with other real-world issues<sup>9,10</sup>. For example, this type of research could enhance our understanding of how animals are dealing with the impacts of climate change, or offer insight into the factors governing behaviours such as habitat selection, mating and foraging. Moreover, understanding how animals move might inspire the

design of robots that can negotiate complex environments. ■

**Andrew A. Biewener** is in the Department of Organismic and Evolutionary Biology, Harvard University, Massachusetts 01730, USA. e-mail: [abiewener@oeb.harvard.edu](mailto:abiewener@oeb.harvard.edu)

1. Wilson, A. M. *et al. Nature* **554**, 183–188 (2018).
2. Hawkes, L. A. *et al. Proc. Natl Acad. Sci. USA* **108**, 9516–9519 (2011).
3. Portugal, S. J. *et al. Nature* **505**, 399–402 (2014).
4. Tsoar, A. *et al. Proc. Natl Acad. Sci. USA* **108**,

E718–E724 (2011).

5. Wittemyer, G., Getz, W. M., Vollrath, F. & Douglas-Hamilton, I. *Behav. Ecol. Sociobiol.* **61**, 1919–1931 (2007).
6. Nagy, M., Ákos, Z., Biro, D. & Vicsek, T. *Nature* **464**, 890–893 (2010).
7. Schaller, G. B. *The Serengeti Lion: A Study of Predator–Prey Relations* (Univ. Chicago Press, 1972).
8. Moore, T. Y., Cooper, K. L., Biewener, A. A. & Vasudevan, R. *Nature Commun.* **8**, 440 (2017).
9. Hussey, N. E. *et al. Science* **348**, 1255642 (2015).
10. Kays, R., Crofoot, M. C., Jetz, W. & Wikelski, M. *Science* **348**, aaa2478 (2015).

This article was published online on 29 January 2018.

## ASTROPHYSICS

# A chirp, a roar and a whisper

In 2017, gravitational waves and electromagnetic radiation were detected from the merger of two stellar remnants called neutron stars. An observational analysis reveals how this radiation was released from the merger. [SEE LETTER P.207](#)

RALPH WIJERS

Last year, scientists reported the coalescence of two astronomical objects known as neutron stars<sup>1</sup>. The event, called GW170817, produced gravitational waves, which had weakened to a faint ‘chirp’ by the time they reached us. In addition, some of the matter in the neutron stars was ejected into space. Moments later, this matter was hit by a powerful jet of material from the merged stars, resulting in a roaring outburst of radiation at all wavelengths<sup>2</sup>. However, despite a flood of data, the process by which this radiation was generated has not been certain. On page 207,

Mooley *et al.*<sup>3</sup> report that GW170817 still whispers to us in radio waves. These signals suggest that the observed radiation came from a relatively slow-moving ‘cocoon’ of matter that was energized by the jet.

The 1993 and 2017 physics Nobel prizes were awarded for the indirect<sup>4,5</sup> and direct<sup>6</sup> detection of gravitational waves, respectively. These studies concerned systems that can be well described using only Einstein’s theory of general relativity. But astrophysics is rarely so simple. For instance, when two neutron stars merge, they produce fireworks — they deform, splash, explode and radiate. Consequently, all the complexities of fluid dynamics, magnetic

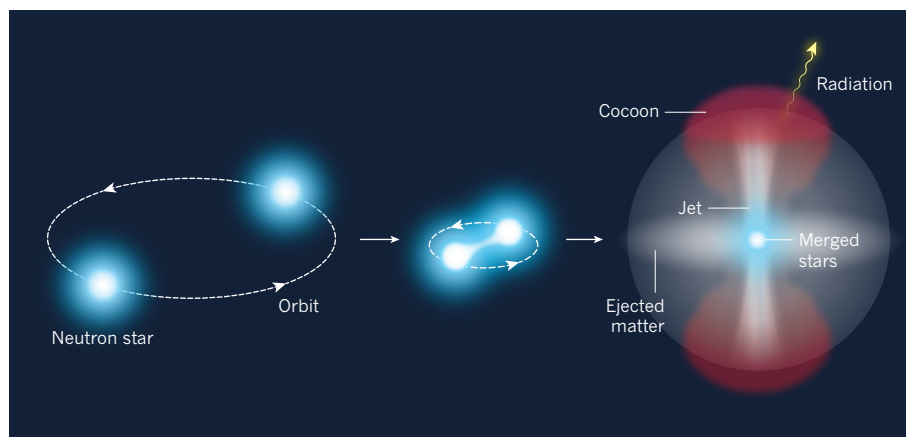
fields, nuclear reactions, particle acceleration and radiation come into play. Astronomers cannot create and tune experiments, but must make do with the messy ones performed by nature.

What astronomers can do, however, is take advantage of two of the biggest revolutions in the field since the invention of the telescope. First, in the twentieth century, astronomy became multi-wavelength: we can now detect radiation across the electromagnetic spectrum (from radio waves to  $\gamma$ -rays). Second, in this century, it became multi-messenger: we can now detect a broad range of emissions — from high-energy cosmic rays and neutrinos to gravitational waves. The discovery of GW170817 demonstrated the full potential of these advances for the first time.

After being alerted to the gravitational-wave signal, astronomers used just about every type of telescope available to try to view the event. As a result, a wide variety of data was obtained, potentially providing enough information to pin down a complete picture of what physically happened when the neutron stars merged. In particular, NASA’s Fermi Gamma-ray Space Telescope detected a flash of  $\gamma$ -rays that had formed within two seconds of the merger<sup>7</sup>. The properties of the flash were consistent with a  $\gamma$ -ray burst (a cosmic explosion long thought to be related to neutron-star mergers), which immediately increased interest in GW170817. However, the exact cause of the  $\gamma$ -ray emission became a matter of debate.

Standard  $\gamma$ -ray bursts can be produced only by a jet — an outflow of material moving at a speed at least 99.9% that of light. But the burst from GW170817 was about 10,000 times weaker than these bursts and seen only because it occurred relatively close to us<sup>7</sup>. Such a weak burst could have come from an off-axis jet (one that was aimed away from us), which would allow only the tiny fraction of light that it emitted sideways to be observed. But it could also have been produced by a comparatively slow-moving cocoon of matter, perhaps travelling at ‘only’ 95% of the speed of light (Fig. 1).

The initial papers<sup>2,8</sup> concluded that both scenarios are possible, and that additional data should allow us to identify which one is correct. Mooley *et al.* now fulfil this promise. They show that although the outburst of radiation



**Figure 1 | Radiation from a neutron-star merger.** A pair of stellar remnants called neutron stars can orbit each other, gradually getting closer, before eventually merging. In 2017, electromagnetic radiation was detected from a neutron-star merger<sup>2</sup>. Mooley *et al.*<sup>3</sup> report evidence for a model that explains how this radiation was generated. In the model, some of the matter in the neutron stars is ejected. This matter is then energized by a powerful jet of material from the merged stars, creating a relatively slow-moving ‘cocoon’ of matter. The cocoon then emits the observed radiation.

from GW170817 is dying down, the intensity of its radio emission is rising — a finding that is difficult to explain using the relatively simple jet models that they consider. The presence of an off-axis jet that breaks free of the surrounding material is not completely excluded, but the cocoon model is more consistent with the observational data.

Establishing the origin of the electromagnetic emission from GW170817 is key to gaining a detailed understanding of the relationship between gravitational-wave events, neutron-star mergers and  $\gamma$ -ray bursts. If a consideration of a greater number of jet and cocoon models than that of Mooley *et al.*, and high-quality simulations of these models, support the authors' conclusions, nature will once again have shown us that the range of phenomena possible is wider than our simplest thinking suggested. If the cocoon model is correct, this probably implies that many more gravitational-wave events have associated  $\gamma$ -ray bursts than was previously thought.

However, Mooley and colleagues' explanation for the burst does not affect our basic understanding of what happens in a neutron-star merger — all of the models considered by the authors have great commonality. For instance, the merged stars are always surrounded by matter that is both inflowing (in the equatorial plane of the merged stars) and outflowing (in all other directions). And a faster and narrower outflow is always driven into this matter along the rotational axis of the merged stars.

The differences are in the precise outcome of the attendant fluid dynamics. How much mass is contained in slow parts of the surrounding matter, and how fast and narrow is the jet? Does the jet break out of the surrounding matter so that it can be seen by us, or does it share its energy with this material, producing a relatively slow and broad explosion?

The next few years will see fierce efforts to address these questions. Gravitational-wave events will be under surveillance by an army of telescopes, to find or exclude a fast jet. And elaborate computer simulations will be used to try to determine what we should expect to happen in neutron-star mergers from a theoretical standpoint. The next handful of well-observed events will bring us much closer to the answers. ■

**Ralph Wijers** is at the Anton Pannekoek Institute for Astronomy, Faculty of Science, University of Amsterdam, 1098 XH Amsterdam, the Netherlands.  
e-mail: ralph.wijers@uva.nl

908–920 (1982).  
6. Abbott, B. P. *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) *Phys. Rev. Lett.* **116**, 061102 (2016).

7. Goldstein, A. *et al.* *Astrophys. J.* **848**, L14 (2017).  
8. Kasiwal, M. M. *et al.* *Science* **358**, 1559–1565 (2017).

## OPTICAL PHYSICS

# Solitons divide and conquer

**An experimental technique allows packets of light called solitons to maintain their shape in all three dimensions as they travel through a material. Such wave packets could find applications in optical information processing.**

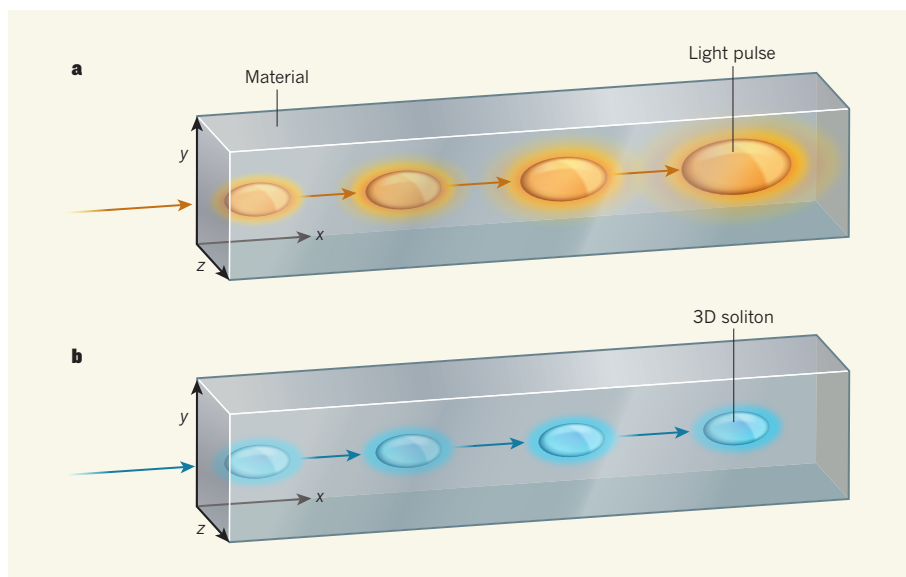
FRANK W. WISE

Waves spread out as they propagate. A familiar example is the broadening of a beam of light. The challenge of overcoming the ubiquitous spreading of waves has motivated scientists for decades, and packets of light waves that retain their shape, known as solitons, have been demonstrated in one and two dimensions<sup>1</sup>. However, it has been extremely difficult to create solitons that are stable in three dimensions. Writing in *Physical Review X*, Lahav *et al.*<sup>2</sup> report an experimental approach that can produce such objects. The work will allow fundamental properties of 3D solitons to be investigated, and could lead to 3D solitons that have technological applications.

A narrow beam of light contains wave

components that propagate in different directions. As the beam travels through a material, these wave components get out of sync, causing the beam to spread out — a process known as diffraction. However, if the beam is powerful enough, the light changes the material's refractive index (a quantity that describes how light propagates in a medium), which in turn affects the beam. In particular, if the beam has a bell-shaped intensity profile, as do most laser beams, the material focuses the beam like a lens. By tuning the beam intensity, this focusing can counteract diffraction to produce a 'self-guided' beam that does not spread out.

In addition to diffraction, a pulsed beam exhibits a broadening effect along its direction of propagation. Each pulse of light contains wave components that have a range of frequencies (colours), and, as a pulse moves



**Figure 1 | Light pulses versus 3D solitons.** **a**, A pulse of light tends to spread out as it propagates through a material (coloured arrows). It broadens along the direction in which it is travelling (the  $x$  axis) as a result of dispersion, whereby components of the pulse that have different frequencies separate. Furthermore, it widens along the perpendicular directions (the  $y$  and  $z$  axes) because of diffraction. **b**, Lahav *et al.*<sup>2</sup> report a technique for producing three-dimensional solitons — packets of light that maintain their geometry as they move through a material.

1. Abbott, B. P. *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) *Phys. Rev. Lett.* **119**, 161101 (2017).
2. Abbott, B. P. *et al.* *Astrophys. J.* **848**, L12 (2017).
3. Mooley, K. P. *et al.* *Nature* **554**, 207–210 (2018).
4. Hulse, R. A. & Taylor, J. H. *Astrophys. J.* **195**, L51–L53 (1975).
5. Taylor, J. H. & Weisberg, J. M. *Astrophys. J.* **253**,

through a material, these components separate — a process called dispersion. There are two types of dispersion: normal and anomalous. In normal dispersion, the low frequencies move faster than the high frequencies ('red leads blue'), whereas in anomalous dispersion, the high frequencies lead the low frequencies ('blue leads red').

However, a high-power beam can cause the change in the material's refractive index to shift the lower frequencies ('red') to the front of the pulse and the higher frequencies ('blue') to the rear. By tuning the beam intensity, the effect of anomalous dispersion can be cancelled out. Furthermore, if a bright beam is turned off and then back on (a dark pulse), the frequency shifts are reversed and normal dispersion can be neutralized.

A 3D soliton, sometimes referred to as a light bullet, is the result of cancelling out diffraction and dispersion simultaneously (Fig. 1). Although such objects exist in theory, they are notoriously unstable. The focusing of the beam by the material must perfectly balance diffraction, and it is extremely difficult to counteract diffraction and dispersion at the same time, because these actions require different beam intensities. Scientists have generated solitons that are stable in two dimensions (one along the direction of propagation and one perpendicular to this direction)<sup>3</sup>, and 3D solitons in a highly structured material (glass patterned with an array of optical devices called waveguides)<sup>4</sup>. But it has not been possible to create 3D solitons in an unstructured

material — which is desirable for studying these objects and for practical applications.

It has been known for more than 20 years that self-guided light beams can be generated in photorefractive materials<sup>5</sup>. These are materials that exhibit a temporary change in refractive index when exposed to a beam of light, as a result of electrons moving through them. The focusing of the beam occurs in such a way that the need for perfect control of the beam intensity is eliminated. Furthermore, bound electrons produce the frequency-shifting refractive index required to eliminate dispersion. There is only one hitch: the material needs to have regions of negative and positive electric charge, but such a charge distribution takes time to establish — longer than the duration of a short light pulse.

Lahav and colleagues' solution was to shine a repetitive string of such pulses into a crystal of the photorefractive material strontium barium niobate, which responded to the power averaged over many pulses to create a self-guided beam. The response of the bound electrons in the crystal then allowed dispersion to be cancelled out in each pulse. The result was a string of 3D 'pulse-train' solitons — so named because the properties of each soliton depend on the solitons that come before it.

The authors used pulses of 800-nanometre wavelength, which meant that the crystal had normal dispersion. As a consequence, although the beam produced was bright, its temporal profile consisted of a dark pulse (see Figure 2 of the paper<sup>2</sup>). By performing similar experiments

at longer wavelengths, for which the crystal has anomalous dispersion, it should be possible to generate bright 3D solitons — one of the major goals in the field of nonlinear optics.

Considering the difficulty in controlling localized 3D wave packets, Lahav and colleagues' results constitute a substantial advance. Interest in localized but non-spreading 3D wave packets extends well beyond optics, to areas as disparate as exotic states of matter known as Bose–Einstein condensates<sup>6</sup> and excitations of substances called ferromagnetic colloids<sup>7</sup>. Furthermore, it should now be possible to investigate how 3D solitons interact when they collide — do they pass right through each other, interact or merge? This information might be useful some day for optical information processing<sup>8</sup>. ■

**Frank W. Wise** is in the School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA.  
e-mail: frank.wise@cornell.edu

1. Kivshar, Y. S. & Agrawal, G. *Optical Solitons: From Fibers to Photonic Crystals* (Academic, 2003).
2. Lahav, O. *et al. Phys. Rev. X* **7**, 041051 (2017).
3. Liu, X., Qian, J. & Wise, F. W. *Phys. Rev. Lett.* **82**, 4631–4634 (1999).
4. Minardi, S. *et al. Phys. Rev. Lett.* **105**, 263901 (2010).
5. Shih, M. *et al. Opt. Lett.* **21**, 324–326 (1996).
6. Strecker, K. E., Partridge, G. B., Truscott, A. G. & Hulet, R. G. *Nature* **417**, 150–153 (2002).
7. Ackerman, P. J. & Smalyukh, I. I. *Nature Mater.* **16**, 426–432 (2017).
8. McLeod, R., Wagner, K. & Blair, S. *Phys. Rev. A* **52**, 3254–3278 (1995).

This article was published online on 5 February 2018.

## MOLECULAR BIOLOGY

# Smoking gun for a rare mutation mechanism

**In 1953, James Watson and Francis Crick proposed that rarely formed isomers of DNA bases cause spontaneous mutations to occur during the copying of DNA. Sixty-five years later, it looks as though they were right. SEE ARTICLE P.195**

MYRON F. GOODMAN

**H**ow do mutations arise when DNA is copied in cells? On page 195, Kimsey *et al.*<sup>1</sup> combine observations of DNA structure with measurements of enzyme kinetics and computational modelling to provide a definitive explanation of a seminal mechanism.

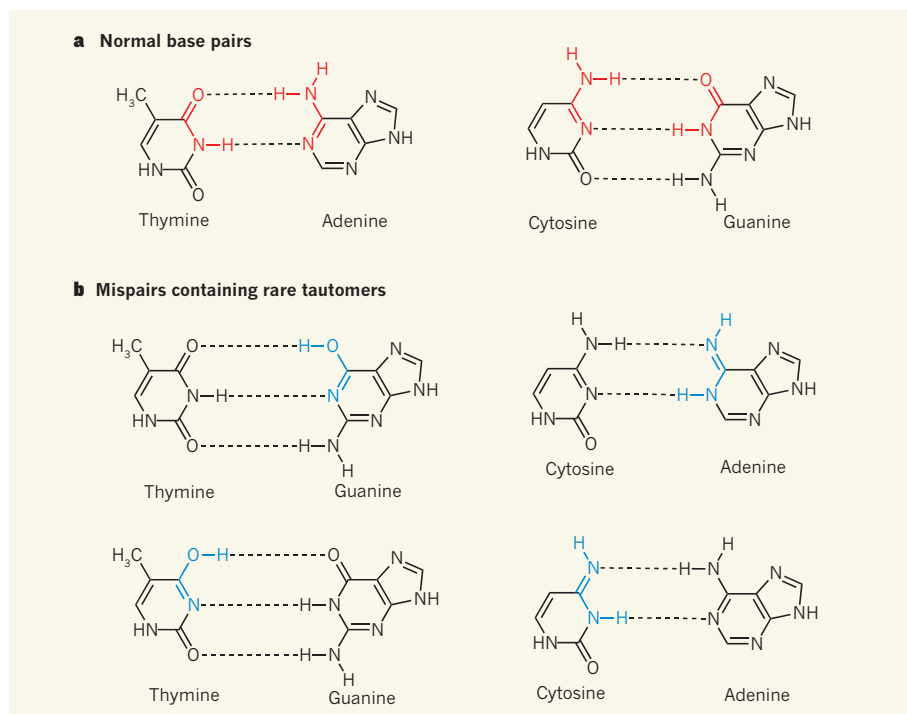
The elucidation of the structure of DNA reported in James Watson and Francis Crick's classic 1953 paper<sup>2</sup> is a monumental piece of work. The key finding was that DNA has a double-helix structure held together by specific interactions between pairs of bases, now known as Watson–Crick pairs: adenine (A)

pairs up with thymine (T), whereas guanine (G) pairs with cytosine (C). The DNA bases exist as tautomers (readily interconvertible pairs of isomers), and the A–T and G–C base pairs contain each base in its predominant tautomeric form (Fig. 1a). The process by which DNA is assembled was not known at the time, but the discovery of base-pairing suggested that the sequence of nucleotides on one strand of a double helix could govern the sequence that was constructed on the complementary strand<sup>2</sup>.

The DNA structure had other far-reaching implications: it suggested a model for how mutations might arise spontaneously as DNA

is made. Watson and Crick proposed<sup>3</sup> that mutations could occur because of "a base occurring very occasionally in one of the less likely tautomeric forms, at the moment when the complementary chain is being formed". In other words, G–T and A–C mismatches could occur if one of the bases is in a disfavoured tautomeric form (Fig. 1b). Such mutations would be easily accommodated because tautomeric mismatches do not distort the helical DNA structure. The disfavoured-tautomer model for spontaneous mutation formation (mutagenesis) was rapidly adopted by biologists and included in textbooks, despite the absence of supporting experimental evidence.

Mispaired structures other than those associated with tautomerization were discovered in the mid-1960s; these included the wobble pairs<sup>4</sup> proposed by Crick, and Hoogsteen pairs<sup>5</sup>. In the mid-1980s, mismatches associated with charged forms of DNA bases were also identified<sup>6–8</sup>. But it wasn't until 2011 that a C–A mismatch associated with a rare tautomer was finally observed in an X-ray crystal structure<sup>9</sup>. The mismatch was formed between bases of nucleotides bound in the active site of DNA polymerase (the enzyme that synthesizes DNA from nucleotides) when DNA synthesis was performed in the presence



**Figure 1 | Base-pair structures in DNA.** **a**, The double-helix structure of DNA is held together by specific interactions (dotted lines) between pairs of bases: adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C). **b**, The DNA bases form rare isomeric structures known as tautomers, which can allow the formation of mispairs; bonds shown in blue are the tautomeric forms of the bonds shown in red in **a**. Kimsey *et al.*<sup>1</sup> have detected tautomeric G-T mispairs in DNA duplexes, and conclude from modelling studies that this explains the frequency with which G-T is misincorporated into DNA during DNA duplication by polymerase enzymes — as proposed<sup>3</sup> by Watson and Crick in 1953.

of manganese(II) ions, which are known to cause mutations. A second X-ray structure<sup>10</sup> reported that year identified an ionized G-T mismatch, also formed between substrates bound by DNA polymerase. In both cases, the mismatched pairs had the same geometry as Watson-Crick pairs.

In Watson and Crick's model for mutagenesis, the rare occurrence of disfavoured tautomeric bases could account for the observed frequency with which DNA polymerases produce mismatches (about one per thousand to one per million base pairs formed<sup>11</sup>). But such tautomers and the associated base pairs were thought to be almost impossible to detect in duplexes. Then, in 2015, 62 years after the mutagenesis model was proposed, researchers from the same group as Kimsey *et al.* reported a tour de force of experimental work: they used nuclear magnetic resonance (NMR) spectroscopy to identify<sup>12</sup> a long-lived wobble G-T structure that was in a dynamic equilibrium with transient, rarely formed G-T mispairs associated with disfavoured tautomers, and with ionized G-T<sup>-</sup> structures, both of which have Watson-Crick geometry.

The first step of the DNA-synthesis process that forms a G-T pair is the binding of dGTP (a G-containing nucleotide) in the polymerase's active site. This is followed by the enzyme's catalytic step, in which the DNA is elongated through incorporation of a new G-T base pair.

Once dGTP is bound in the active site, the base pair formed between dGTP and T on the complementary strand assumes a distorted wobble conformation, but seemingly cannot make the conformational transition needed for the catalytic step<sup>10</sup>.

Kimsey and colleagues' current study goes straight to the heart of the mutagenesis model by integrating structural analysis of G-T base pairs in duplexes with measurements of the kinetics of DNA polymerase reactions and computer modelling to show that tautomerism does indeed account for the misincorporation of base pairs. To ensure efficient catalysis, DNA polymerases require optimal geometrical alignment of nucleotide substrates with amino-acid residues in their active site<sup>13,14</sup>. Such alignment can occur when G-T adopts one of its Watson-Crick-like structures (one of the disfavoured tautomeric forms, or the ionized structure<sup>15</sup>). Kimsey *et al.* deduced from their studies that, at neutral pH, at least 99% of G-T misincorporation is attributable to the formation of G-T tautomers — rather than of the ionized structure — from an initially bound G-T wobble pair.

By successfully identifying a role for the disfavoured tautomeric forms of G-T in base-pair misincorporation, Kimsey and colleagues have solved half of the mystery of spontaneous mutagenesis. A solution for the other half now requires the disfavoured

tautomeric forms of C-A to be characterized in duplexes and correlated with the rate of C-A misincorporation. So far, NMR and X-ray data have identified only charged C-A<sup>+</sup> wobble structures in a DNA duplex<sup>7,8</sup>.

A related challenge would be to establish the mechanism by which 2-aminopurine, a base analogous to both adenine and guanine, induces mutagenesis. For example, 2-aminopurine is a potent mutagen of the virus bacteriophage T4, for which it increases the frequency of A-T to G-C mutations (and of the reverse G-C to A-T mutations) to 10–50 times the frequency of spontaneous mutation levels<sup>16</sup>. If 2-aminopurine was found to undergo a tautomeric shift much more frequently than A, it would implicate tautomerization in the mechanism, and thus provide the icing on the cake for the tautomerization model of mutagenesis. ■

**Myron F. Goodman** is in the Molecular and Computational Biology Program, Departments of Biological Science and Chemistry, University of Southern California, Los Angeles, Los Angeles, California 90089, USA.

e-mail: mgoodman@usc.edu

- Kimsey, I. J. *et al.* *Nature* **554**, 195–201 (2018).
- Watson, J. D. & Crick, F. H. C. *Nature* **171**, 737–738 (1953).
- Watson, J. D. & Crick, F. H. C. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131 (1953).
- Crick, F. H. C. *J. Mol. Biol.* **19**, 548–555 (1966).
- Hoogsteen, K. *Acta Crystallogr.* **16**, 907–916 (1963).
- Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F. & Fazakerley, G. V. *J. Biol. Chem.* **263**, 14794–14801 (1988).
- Sowers, L. C., Fazakerley, G. V., Kim, H., Dalton, L. & Goodman, M. F. *Biochemistry* **25**, 3983–3988 (1986).
- Hunter, W. N., Brown, T., Anand, N. N. & Kennard, O. *Nature* **320**, 552–555 (1986).
- Wang, W., Hellinga, H. W. & Beese, L. S. *Proc. Natl Acad. Sci. USA* **108**, 17644–17648 (2011).
- Bebenek, K., Pedersen, L. C. & Kunkel, T. A. *Proc. Natl Acad. Sci. USA* **108**, 1862–1867 (2011).
- Kunkel, T. A. & Bebenek, K. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
- Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. *Nature* **519**, 315–320 (2015).
- Echols, H. & Goodman, M. F. *Annu. Rev. Biochem.* **60**, 477–511 (1991).
- Tsai, Y.-C. & Johnson, K. A. *Biochemistry* **45**, 9675–9687 (2006).
- Yu, H., Eritja, R., Bloom, L. B. & Goodman, M. F. *J. Biol. Chem.* **268**, 15935–15943 (1993).
- Drake, J. W. & Allen, E. F. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 339–344 (1968).

This article was published online on 31 January 2018.

## CORRECTION

The News & Views 'Strategy for making safer opioids bolstered' by Susruta Majumdar and Lakshmi A. Devi (*Nature* **553**, 286–288; 2018) incorrectly stated that more than 100,000 adults suffer from chronic pain in the United States. The correct figure is more than 100 million adults.

# Biomechanics of predator–prey arms race in lion, zebra, cheetah and impala

Alan M. Wilson<sup>1</sup>, Tatjana Y. Hubel<sup>1</sup>, Simon D. Wilshin<sup>1</sup>, John C. Lowe<sup>1</sup>, Maja Lorenc<sup>1</sup>, Oliver P. Dewhurst<sup>1</sup>, Hattie L. A. Bartlam–Brooks<sup>1</sup>, Rebecca Diack<sup>1</sup>, Emily Bennitt<sup>2</sup>, Krystyna A. Golabek<sup>3</sup>, Roger C. Woledge<sup>1‡</sup>, J. Weldon McNutt<sup>3</sup>, Nancy A. Curtin<sup>1</sup> & Timothy G. West<sup>1</sup>

**The fastest and most manoeuvrable terrestrial animals are found in savannah habitats, where predators chase and capture running prey. Hunt outcome and success rate are critical to survival, so both predator and prey should evolve to be faster and/or more manoeuvrable. Here we compare locomotor characteristics in two pursuit predator–prey pairs, lion–zebra and cheetah–impala, in their natural savannah habitat in Botswana. We show that although cheetahs and impalas were universally more athletic than lions and zebras in terms of speed, acceleration and turning, within each predator–prey pair, the predators had 20% higher muscle fibre power than prey, 37% greater acceleration and 72% greater deceleration capacity than their prey. We simulated hunt dynamics with these data and showed that hunts at lower speeds enable prey to use their maximum manoeuvring capacity and favour prey survival, and that the predator needs to be more athletic than its prey to sustain a viable success rate.**

In a chase, the prey animal can select its speed and the timing of acceleration, deceleration and turns, whereas a predator in pursuit must predict or respond to the trajectory of its prey to enable interception and capture<sup>1–3</sup>. The prey should make its movements unpredictable to the predator while generally using tactics that minimize the chance the predator has of catching it<sup>4</sup>. Therefore, although the optimum avoidance strategy might be, for instance, to perform a maximum-rate turn away from the predator, using this strategy consistently would enable the predator to anticipate that manoeuvre<sup>5</sup>. If a dominant evolutionary pressure on the locomotor system is predation success or evasion, then predator and associated prey should display similar high levels of athleticism<sup>6–9</sup> distinguished by the specific adaptations necessary to enable capture (predators) or evade capture (prey)<sup>10</sup>. We hypothesize that predators are consistently more athletic than their prey so that they can manoeuvre and change speed to respond to the unpredictable tactics of the prey animal.

We studied two predator–prey pairs found on the southern African savannah, where a simple high-speed manoeuvring pursuit in open terrain is a commonly used hunting technique: cheetah, *Acinonyx jubatus*, and impala, *Aepyceros melampus*<sup>11</sup>, which are similar in size (50–70 kg compared to 50–60 kg, respectively; Methods), and the substantially larger lion, *Panthera leo*, and zebra, *Equus quagga*<sup>12</sup> (120–240 kg compared to 320 kg) (Fig. 1a–d, f).

We evaluated five metrics. First, locomotor muscle maximum power output and contraction speed, which is assumed to be critical for speed, acceleration and turning performance<sup>13,14</sup>. Second, animal acceleration and deceleration (in the direction of travel). This combines muscle power and volume with factors that include grip, body shape<sup>15</sup> and the anatomical arrangement of muscles<sup>16</sup>. Third, the highest speed commonly used by each species and the actual top speed recorded<sup>9</sup>. Fourth, animal turning performance (centripetal or lateral acceleration and heading rate), which can be limited by grip<sup>17</sup>, leg strength<sup>18</sup> and muscle power<sup>19,20</sup>. Finally, stride frequency, because there is only one opportunity per stride for the legs to apply impulses to change speed and direction in a hunt<sup>21</sup>.

Locomotion data were collected from free-living wild animals undertaking high-speed runs in northern Botswana using our own design of Global Positioning System and Inertial Measurement Unit collars<sup>9</sup> (GPS-IMU; Methods, Fig. 1 and Extended Data Fig. 1a). We collected velocity and acceleration data from 23,871 strides from 520 runs of five cheetahs, 22,491 strides from 515 runs of seven impalas, 111,110 strides from 2,726 runs of nine lions and 64,952 strides from 1,801 runs of seven zebras (Extended Data Fig. 1f). Muscle biopsies were collected from the biceps femoris, a major propulsive muscle in the hind leg (Methods).

## Muscle fibre power in predator and prey

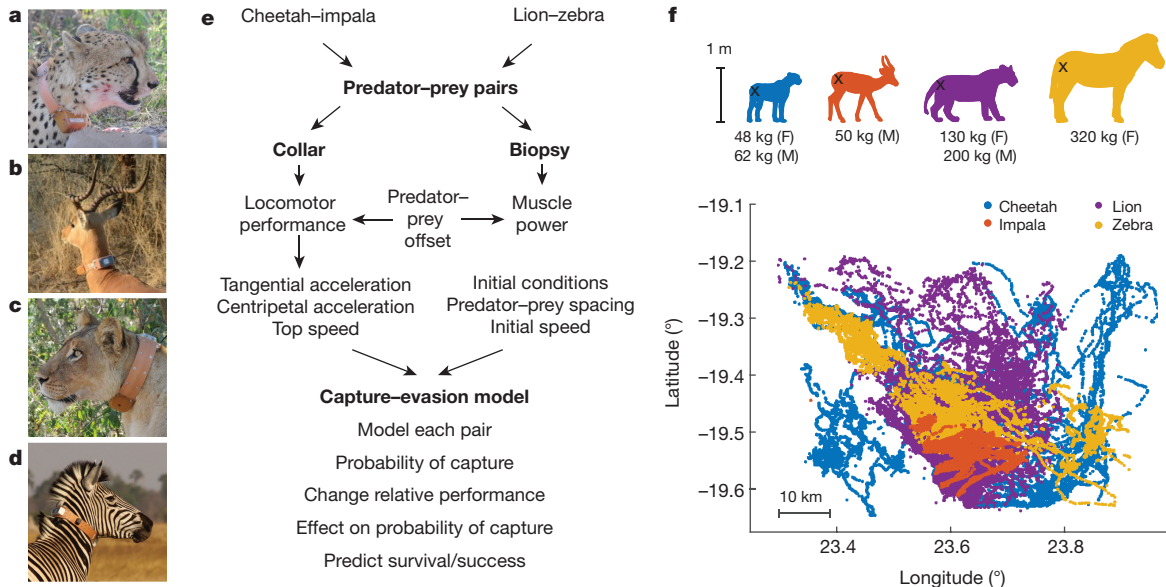
Muscle biopsies were skinned, placed in a trehalose–glycerol mixture, frozen in liquid nitrogen in the field and transported to the United Kingdom. Peak power, velocity and stress at peak power and maximum isometric stress were determined at 25 °C for single, skinned fibres (Fig. 2a–f). Maximum power and associated velocity and stress were then calculated (Methods).

Complete measurements were made on 37 individual skinned fibres from six cheetahs, 30 fibres from five impalas, 50 fibres from eight lions and 57 fibres from eight zebras. There was a distinct subpopulation of ‘low-performance’ fibres (twelve fibres from zebra, eight fibres from lions, three fibres from cheetahs and three fibres from impalas; Fig. 2d–f and Supplementary Data) with a velocity at peak power that was below 1.35 lengths s<sup>−1</sup> and a lower peak power (Fig. 2c and Extended Data Fig. 2g), which were either myosin heavy chain (MHC) type-I (11 of 19 fibres tested) or type-II (8 fibres) (Methods).

Linear mixed-effects models were fitted for peak power, velocity and stress at peak power and isometric stress with a factor distinguishing predator and prey, including the interaction of this factor with a categorical variable called ‘fibre performance classification’. Within the factor distinguishing predator and prey, we nested a random effect by subject and fibre. The residuals of this model exhibited heteroscedasticity and so the variance of the error term was allowed to vary by subject and performance classification. Power in the high-performance fibres

<sup>1</sup>Structure & Motion Lab, Royal Veterinary College, University of London, Hatfield AL9 7TA, UK. <sup>2</sup>Okavango Research Institute, University of Botswana, Maun, Botswana. <sup>3</sup>Botswana Predator Conservation Trust, Private Bag 13, Maun, Botswana.

‡Deceased.



**Figure 1 | Overview.** **a–d**, The four species in the study wearing the collars; the release mechanism is shown in **d**. **e**, Flowchart summarizing study. **f**, Relative animal size and biopsy location (black cross on animal) along with collar GPS position data for the four species showing range overlap. F, female; M, male.

was 20% greater in the predator group than in the prey group (effect size =  $20.0 \text{ W kg}^{-1}$ ,  $z = -3.46$ ,  $P = 0.001$ ). The difference was similar in both pairings. However, the effect was significant only in the lion–zebra pair (effect size =  $20.0 \text{ W kg}^{-1}$ , 20%,  $z = 2.56$ ,  $P = 0.039$ ), but not in the cheetah–impala pair (effect size =  $18.9 \text{ W kg}^{-1}$ , 19%,  $z = 2.04$ ,  $P = 0.15$ ). The peak-specific powers were very similar in the two predator species, and lower but very similar in the two prey species (mean power of high-performing fibres  $\pm$  standard error in  $\text{W kg}^{-1}$ : cheetahs,  $106.7 \pm 4.6$ ; lions,  $108.1 \pm 4.2$ ; impalas,  $88.3 \pm 5.2$ ; zebras,  $88.4 \pm 4.1$ ).

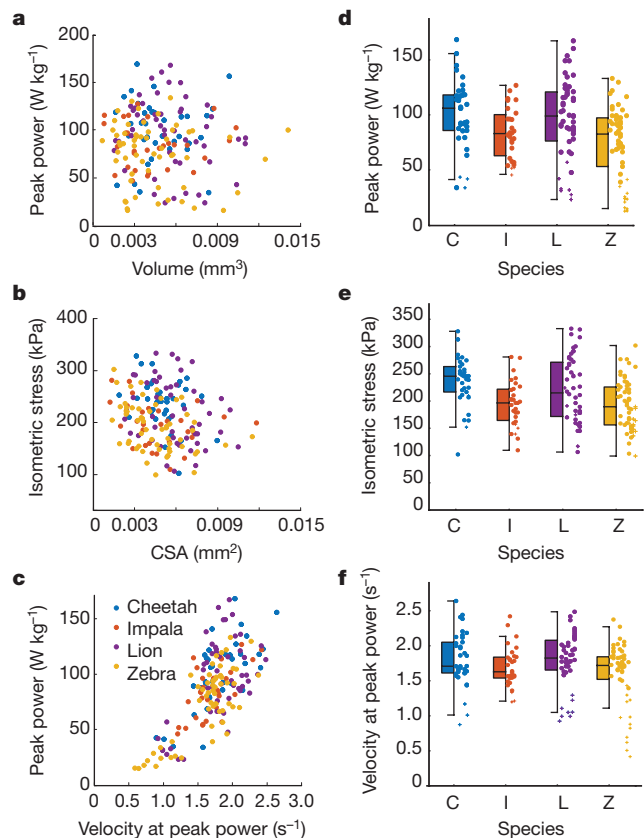
No significant differences between predators and prey were detected for the velocity at peak power (effect size  $0.096 \text{ lengths s}^{-1}$ ,  $z = -1.77$ ,  $P = 0.15$ ) or stress at peak power (effect size =  $7.2 \text{ kPa}$ ,  $z = -2.05$ ,  $P = 0.075$ ) for high-performance fibres. Isometric stress was higher in predators (effect size =  $33.4 \text{ kPa}$ ,  $z = -2.87$ ,  $P = 0.008$ ; Extended Data Fig. 2).

The values reported here are comparable to data for skinned fibres from wild rabbits at  $25^\circ\text{C}$ <sup>22</sup>, but are high compared to published values for skinned fibres from large animals<sup>23,24</sup>. Muscle power is highly temperature dependent<sup>25</sup> and a temperature coefficient ( $Q_{10}$ ; the ratio of the increase in rate with a temperature increase of  $10^\circ\text{C}$ ) of 2.3 is appropriate<sup>26</sup>, which predicts *in vivo* muscle power (all fibres, Extended Data Fig. 2i) of 232 (prey) and 292 (predators)  $\text{W kg}^{-1}$  at a body temperature of  $38^\circ\text{C}$ .

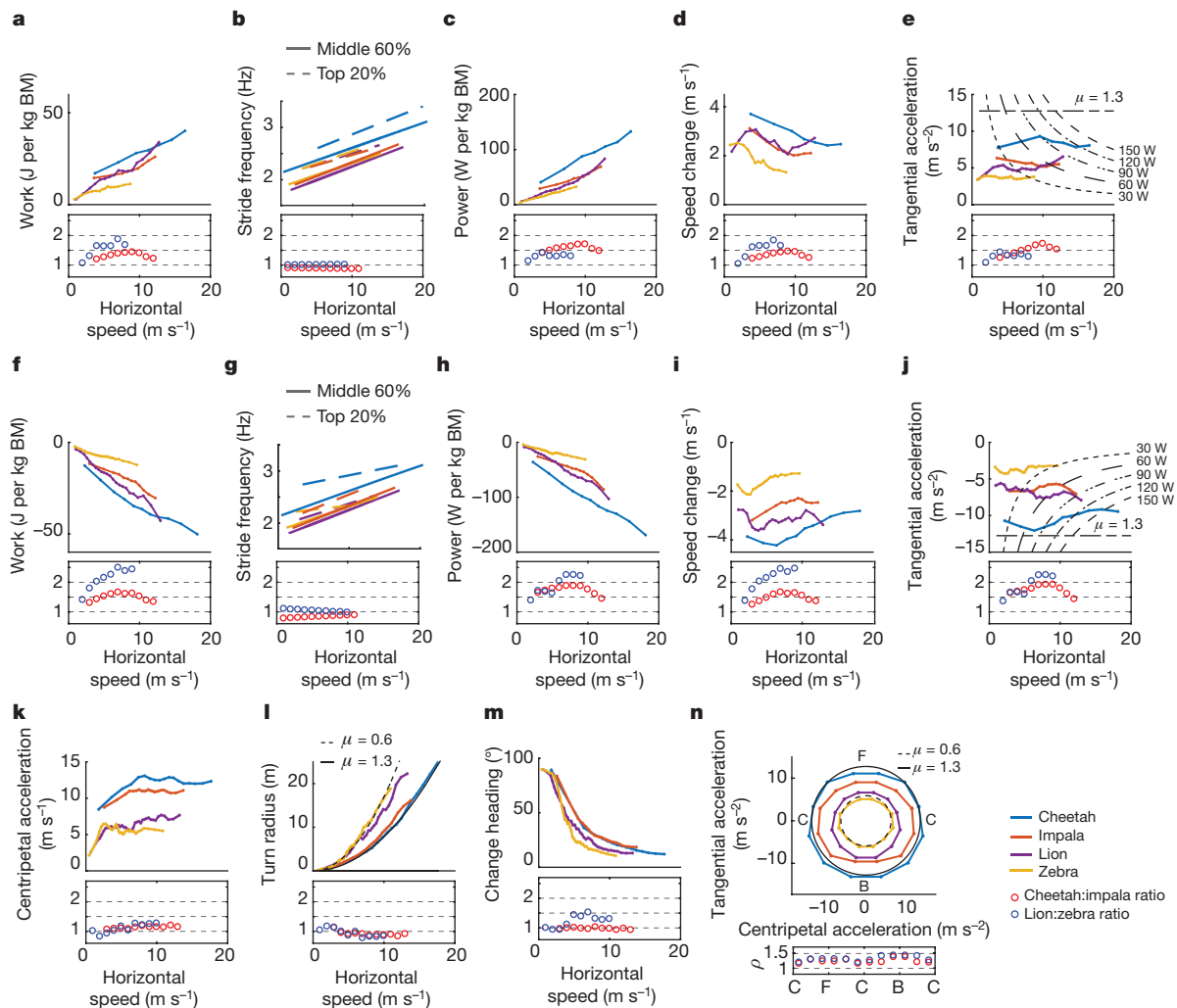
Slower myosin types and muscle fibres are inherently more economical<sup>23,25,27</sup>, thus slower fibres confer advantages<sup>25</sup>, and the fast versus slow distribution of fibres reflects the opposing pressures of predation (avoidance) on one side and food and water supply, ranging distance and environmental conditions on the other<sup>25,28</sup>. This may partly explain why prey species have lower-power muscle fibres<sup>25</sup>. Therefore, muscles of desert specialists at risk of dehydration and/or starvation<sup>29</sup>, such as camels, vicuñas and Arabian oryx, would be predicted to be biased towards economy<sup>25</sup>. Selection pressure for greater performance or economy could change fibre type distributions or muscle characteristics within a few generations—much more rapid than for changes in myosin contractile speed.

### Speed and acceleration of predators and prey

Stride timing and therefore frequency was derived from collar acceleration data<sup>9</sup>. Stride speed and accelerations were averaged over each stride; change in speed is calculated as the difference in speed between



**Figure 2 | Muscle contraction mechanics.** **a**, Peak power compared to volume. **b**, Maximum isometric stress compared to cross sectional area. **c**, Peak power versus velocity at peak power, each point is a muscle fibre. **d–f**, Box plots show variations in power (**d**), isometric stress (**e**) and optimal shortening velocity (**f**) across the four species, with each fibre represented as a dot and the low-performance fibres as a plus symbol. Data from each individual are shown in a separate vertical column. Line indicates the median, box shows the interquartile range (IQR) and the whiskers are  $1.5 \times \text{IQR}$ . Data are from 37 fibres from six cheetahs (C), 30 fibres from five impalas (I), 50 fibres from eight lions (L) and 57 fibres from eight zebras (Z).



**Figure 3 | Locomotor performance based on stride parameters.** All values are averaged per stride or represent the change over a stride and where appropriate are per kg body mass (BM). **a–e**, Accelerating strides. **a**, Positive net work performed in each stride. **b**, Stride frequency, mean of 20% highest power strides and middle 60% of power strides. **c**, Average mass-specific power per stride. **d**, Increase in speed per stride. **e**, Tangential (forward) acceleration with the curved lines representing a stride mean power of 30, 60, 90, 120 and 150  $\text{W kg}^{-1}$  with a limit line for a coefficient of friction ( $\mu$ ) of 1.3. **f–j**, as **a–e**, but for decelerating strides. **k–n**, Turning. **k**, Centripetal (lateral) acceleration. **l**, The relationship between speed and turn radius with limit lines for  $\mu = 0.6$  and  $\mu = 1.3$ .

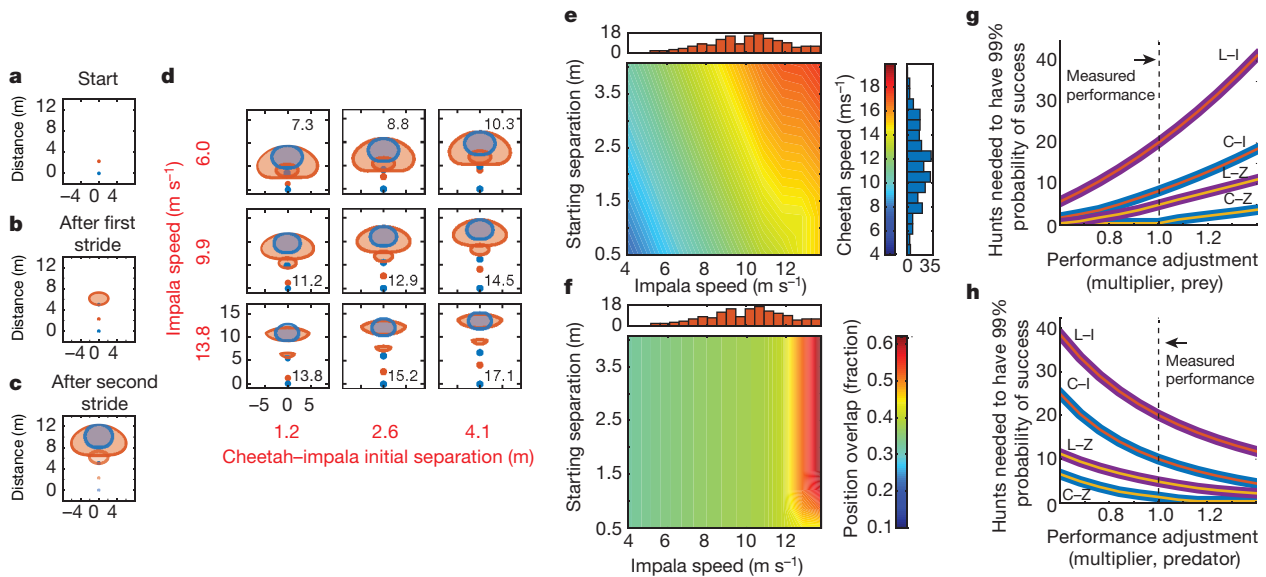
**m**, Change in heading compared to horizontal speed. **n**, Tangential compared to centripetal acceleration with  $\mu$  limits as for **l**. **n**, F, pure forward acceleration; B, deceleration; C, centripetal acceleration;  $\rho$ , polar coordinate. In each panel one line per species is shown, which (except in **b**, **g**) represents the 98th percentile for data in speed bins (each bin contains 400 data points therefore bin width varies). At the bottom of each panel, the ratio of that parameter for cheetah–impala (red circle) and lion–zebra (blue circle) is given for each speed bin, same x axis. Dataset comprised 7,509 strides for 520 runs from five cheetahs; 8,884 strides for 515 runs from seven impalas; 15,947 strides for 2,726 runs from nine lions and 14,089 strides for 1,801 runs from seven zebras.

two consecutive strides, work per stride is the change in mass-specific net horizontal kinetic energy and power per stride is the work per stride divided by stride duration. Change in heading is the angle between two consecutive stride velocity vectors<sup>9</sup>.

Differences in the frequency of maximum effort manoeuvring between predators and prey (since predators hunt often and prey are rarely hunted) would manifest in different tails for the distributions of accelerations for each species. The predator species will have relatively heavy tails, that is, higher kurtosis, as more of their observed behaviours are associated with rapid accelerations, whereas the more sedentary (or at least steadily moving) prey have fewer such observations. Steady-state strides were removed by including a threshold on acceleration with the threshold determined for each species by the kurtosis of these distributions, which resulted in a similar distribution for all species (Methods and Extended Data Fig. 3a–c). Qualitatively, the distributions for predators and for prey are similar and the 98% percentile approximates the limit of the distribution in a reasonably consistent manner across runs of all lengths and tortuosity (Extended Data Fig. 4).

Stride parameters were grouped into non-uniform speed bins with 400 data points in each and the 98th percentile of the distribution was determined for each bin (except for stride frequency, for which data were further subgrouped on the basis of acceleration performance and a linear regression was performed on each subgroup (Methods)). The uppermost bin with fewer than 400 data points was ignored. The 98th percentile was chosen to account for the different numbers of strides in different species and to exclude occasional extreme values<sup>9</sup> (Extended Data Figs 5, 6). The cheetah–impala pair was more athletic than the zebra–lion pair for every metric (Extended Data Fig. 7).

Predator and prey were compared using a linear model (Methods) and test statistics computed under the null hypothesis that predator and prey are drawn from the same distribution, except for stride frequency, for which, because of species pairing differences, predator and prey pairs were compared individually. The ratio of the maximum observed performance for cheetah–impala then lion–zebra, along with the results of the test comparing predator and prey across species, are as follows: predators were 50% and 24% superior at acceleration



**Figure 4 | Output of the model of predator–prey interaction.** **a**, Predator (blue) and prey (red) with initial (upward) velocity and separation. **b**, After one stride the prey can accelerate to anywhere in the red ellipse. Predator velocity remains unchanged as there is no prey acceleration in the previous stride. The initial positions are shown. **c**, The red ellipse perimeter is the area that prey can reach after two strides at the chosen maximum acceleration. The blue ellipse represents the locations the predator can occupy (responding to the prey acceleration observed in first stride). The area of the prey ellipse that is covered by the predator ellipse is defined as the probability of capture. **d**, As in **c** for different initial conditions. Rows are different initial impala speeds, values in red to the left. Columns are different initial separations with values in red below each column. Scale for all instances in the bottom left plot are in metres. The black numbers

( $z = 3.15$ ,  $P = 0.0016$ ); 73% and 70% better at deceleration ( $z = -6.61$ ,  $P < 0.0001$ ); 100% and 89% more powerful during maximal acceleration ( $z = 3.87$ ,  $P = 0.0001$ ); and 100% and 122% more powerful during maximal deceleration ( $z = -8.07$ ,  $P < 0.0001$ ). Stride frequency was higher for cheetahs than for impalas ( $z = 3.69$ ,  $P < 0.001$ ) and lower for lions than for zebras ( $z = -2.31$ ,  $P = 0.041$ ). Across all species, stride frequency at  $8 \text{ m s}^{-1}$  was 6% higher during acceleration ( $P = 0.0018$ ) and 5% higher during deceleration ( $P < 0.001$ ) than during steady speed locomotion, as determined by post hoc tests on the linear model.

The 98th percentile of speed was  $19.9 \text{ m s}^{-1}$  for cheetahs,  $13.8 \text{ m s}^{-1}$  for impalas,  $13.9 \text{ m s}^{-1}$  for lions and  $10.6 \text{ m s}^{-1}$  for zebras. This was 84, 78, 67 and 77%, respectively, of the maximum achieved by the third fastest individual of each species, which was  $23.8 \text{ m s}^{-1}$  for cheetahs,  $17.7 \text{ m s}^{-1}$  for impalas,  $20.6 \text{ m s}^{-1}$  for lions and  $13.8 \text{ m s}^{-1}$  for zebras. Therefore, predators were faster than their prey and all species rarely approached their maximum recorded speed (Extended Data Fig. 5).

### Turning performance of predators and prey

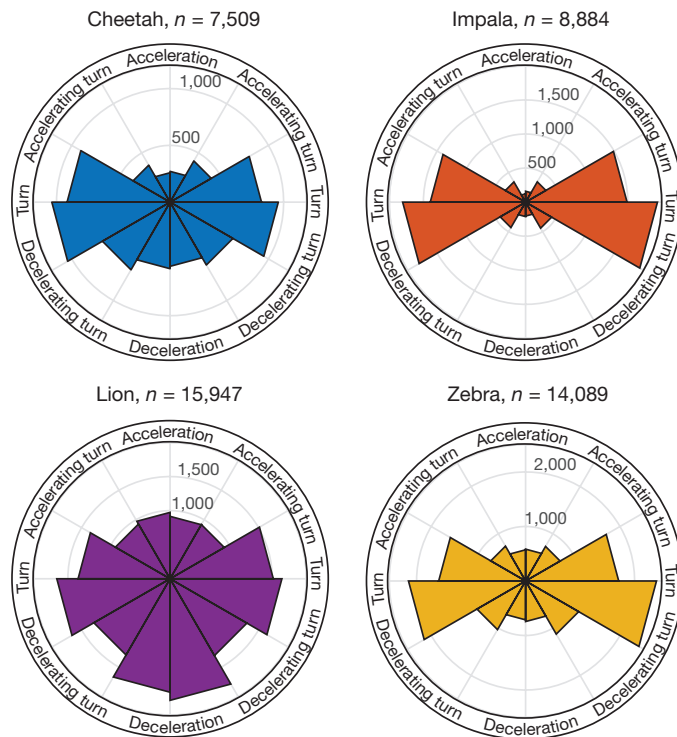
When turning, predators were only slightly superior to prey ( $z = 2.93$ ,  $P = 0.0034$ ): cheetah–impala 15%, lion–zebra 10% (Fig. 3k–n). Turning does not require a change in body kinetic energy, but a centripetal acceleration of  $13 \text{ m s}^{-2}$  results in a 66% increase in effective weight<sup>18</sup> and the limbs must shorten and extend in the presence of these higher axial forces. This length change can be delivered by passive elastic structures within the limb<sup>30,31</sup>, but any associated muscles must deliver higher forces at that contraction velocity (equating to a higher power requirement)<sup>19,20</sup>. Reduced centripetal acceleration at high speed would indicate a muscle power limit rather than a grip limit for that activity<sup>17,19</sup>, however, we found no such evidence for a power limit<sup>18</sup> at these submaximal speeds.

are the initial (optimized for maximum success) cheetah speeds in  $\text{m s}^{-1}$ . **e**, Optimum cheetah speed to maximize overlap (colour key on the right) as a function of impala speed (x axis) and starting separation (y axis). Histogram at the top shows actual impala speeds at first turn of 10 degrees or more (x axis as the main plot) and the vertical histogram shows the distribution of actual cheetah speed at the first turn (scale as heat bar). **f**, Proportional overlap (capture probability) as a function of initial speed of impalas and starting separation. **g**, **h**, Number of hunts required to have a 99% chance of prey capture for different performance levels of prey and predator. Labels and line colours indicate species pairings for each line. Actual accelerations multiplied by the x axis value; the simulation was run at  $8.75 \text{ m s}^{-1}$  initial prey speed and initial predator–prey spacing of 2 m.

Figure 3n summarizes the capacity for maximum acceleration in any direction, relative to the track of the animal. It shows that these predators outperform their prey most markedly during deceleration (bottom) and less so during forward acceleration (top) and turning (sides). No species showed highest levels of tangential and centripetal acceleration in the same stride; the lines are elliptical, which supports a grip-type limit (as horizontal accelerations should vector sum to a limit value). Forward acceleration performance was maintained by all four species at the fastest speeds commonly used (Fig. 3e, j). Power requirements for forward acceleration increase with speed (Fig. 3c, h and Extended Data Fig. 5), because power is the product of speed and acceleration, and if maximal acceleration was lower at the highest speeds, this would indicate a potential power constraint<sup>15</sup>. A reduction in manoeuvrability would result in an animal's trajectory being more predictable, which would be disadvantageous for both predator and prey.

### Species differences in experimental data

Much of the difference observed between predator and prey could be attributed to predators having proportionally more muscle and/or higher muscle power (Fig. 2d), but that does not provide an explanation for the large differences that were observed between lion and cheetah and between zebra and impala (Extended Data Fig. 7). Hind limb muscle fraction of total body mass is fairly consistent across species: 17.5–19.8% (Extended Data Table 1), so muscle peak power should define the acceleration capacity of the whole animal at moderate to high speeds<sup>9,13,15</sup>. Athletic wild animals are, however, likely to be proportionally more muscular than the mostly sedentary domesticated animals contributing to Extended Data Table 1 and spinal, trunk and forelimb muscle will also contribute to acceleration. The predicted *in vivo* muscle powers of 232–292  $\text{W kg}^{-1}$  are concomitant with the upper, but not lower, limit of observed whole animal powers of 30–120  $\text{W kg}^{-1}$  (Fig. 3e).



**Figure 5 | Orientation of direction of acceleration for each stride grouped by species.** Circular histogram of frequency/direction of the acceleration vector for each stride (steady state strides were removed) binned in twelve 30-degree sectors. Upwards is forward acceleration, down is deceleration, left and right are turns in that direction. Height of each bin from the centre indicates the number of strides. The dataset comprised 7,509 strides for 520 runs from five cheetahs; 8,884 strides for 515 runs from seven impalas; 15,947 strides for 2,726 runs from nine lions and 14,089 strides for 1,801 runs from seven zebras.

Carnivores hunt with empty stomachs, whereas prey carry the mass of rumen (impala) or hind gut (zebra) contents, which will impinge on any performance that is dependent on muscle power or strength (as would pregnancy). The differences within the predator and within the prey species may reflect that muscles are arranged for different roles, for example, for economical walking versus for acceleration and hunting or fighting<sup>25,32,33</sup>, but without contextual anatomical data, this is only speculation and the differences are too large to simply be attributed to scaling due to animal size (Extended Data Table 2a). Foot design and grip may also have a role<sup>34</sup>. Behavioural factors cannot be ruled out, but our data indicate that the highest values were captured (Extended Data Figs 5, 6).

### Capture–evasion model description and predictions

A pursuit predator uses a combination of stealth and speed to get close to its prey<sup>12</sup> and then the prey evades capture by manoeuvring<sup>1,4</sup>, while the predator attempts to intercept it. The interaction has been approached analytically or numerically for continuous processes<sup>35,36</sup> (for example, air combat manoeuvring), but modelling the probability that the predator and prey arrive at the same location becomes increasingly complex to solve when treated as a discrete process.

In our model, predator and prey were able to accelerate in any direction up to their experimentally derived maximum during each stride, so they could go anywhere on the boundary of an approximately elliptical area that grew with the subsequent stride (Fig. 4a–c). The predator responds to the acceleration of its prey in the preceding stride and we modelled the range of initial conditions for which the predator could catch its prey within two strides. The acceleration limits for each species and direction (impulses per stride) were the observed 98% values of

centripetal, positive and negative tangential acceleration divided by the stride frequency at that speed (Fig. 3b, e, g, j). The elliptical area prevented simultaneous maximal centripetal and tangential accelerations (Fig. 3n). At higher speeds, acceleration and therefore manoeuvring were curtailed as the applied impulses could not cause the animal to exceed the 98% maximum speed observed for each species.

For the prey, the accelerations at the start of the first and second stride were the possible accelerations up to maximum (Extended Data Table 2b) in any direction. The predator had zero acceleration in the first stride so its initial velocity determined its subsequent position and it could accelerate in any direction in the second stride (reacting to the previous prey acceleration). The area reached by the predator was increased by a semi-circular region with a size of half the predator body length to account for the physical size of the predator. We define capture probability as the fraction of the elliptical area of the prey that is covered by the elliptical area of the predator after two strides.

We plotted the feasible range of initial prey speeds and predator–prey spacings for capture after two strides (Fig. 4d and Extended Data Fig. 8) and then optimized the initial predator speed for each condition to maximize the overlap in position between the predator and the prey after two strides. The predator–prey spacing at the beginning of the simulation represents less than a stride length in all cases (code used for the model written in Python can be found in the Supplementary Information).

The model shows that the prey should avoid the predator by turning (lateral acceleration), rather than attempting to increase separation by travelling as fast as possible (Fig. 4d). If the prey is moving fast and the predator is close (Fig. 4d, bottom left), its best option requires rapid deceleration and turning, whereas turning alone becomes more beneficial if the predator is further away (and therefore closing at higher relative speed, Fig. 4d, bottom right). High prey speeds result in high capture probabilities (Fig. 4f), because the prey cannot accelerate forwards with or without turning, making its tactics highly predictable (captured by optimization of predator speed for overlap), whereas a slow moving prey (Fig. 4e, f, left) has a wider variety of escape options and is therefore less predictable. Predator and prey indeed use moderate speeds (Fig. 4e and Extended Data Fig. 8).

The predator has the highest chance of success if it is travelling only slightly faster than the prey, which enables it to reach many of the locations the prey can move to across a broad range of starting speeds (the objective function for the optimization, relative capture area, is very flat in this region), and its advantage increases with higher prey speeds. This is reflected by the observed predator speeds (Fig. 4e and Extended Data Fig. 8).

Figure 5 shows that all species often execute a constant speed turn, but that it is rare for either of the herbivore species to accelerate or decelerate, whereas predators (especially lions) often undertake deceleration strides, either in isolation or in combination with a turn. The preferred accelerations fit with the prey using optimum escape strategies predicted by the non-overlapping areas in Fig. 4d and tactics for which they perform similar to the performance of predators (turning) rather than those for which they are outperformed (tangential acceleration and deceleration). With the same lateral acceleration, a prey that is moving more slowly than a converging faster-moving predator will have an advantageously tighter turn, because the radius is equal to  $v^2/\text{lateral acceleration}$ . Commonly observed predator decelerations are concomitant with a faster-moving closing predator. More than one repetition of the modelled two-stride scenario can occur within a single pursuit<sup>9</sup>—and the overlap-derived success rates are similar to those observed for animals when hunting in the wild<sup>9,12,37</sup>.

### Effect of athleticism on hunting success rate

We adjusted the acceleration capacity of the predator or prey and reran the simulation to obtain capture probabilities for animals of greater or lesser athleticism. Unsurprisingly, increased predator performance is beneficial, reducing the number of hunts needed to capture prey

(Fig. 4g, h). Owing to the power relationship underlying Fig. 4g, h, curves steepen when the predator is below 0.8 of its actual performance (Fig. 4h), which would tend towards an unsustainably low success rate (ignoring other determinants of hunt outcome). Such a reduction could be the result of an injury or ageing, with greatest consequences for solitary animals. The data also provide insight into preferred prey and hunting style: the predicted low success rate for lions hunting impala (Fig. 4g, h) is supported by the observation that lions capture impala opportunistically rather than in an open pursuit. African wild dogs hunt impala<sup>37</sup>, but are less athletic than cheetahs<sup>37</sup>. Applying the model to a single African wild dog hunting an impala<sup>37</sup> predicts a success rate of 8.2%, which is lower than the actual success rate of 15.5%<sup>37</sup>. This would concur with African wild dogs capturing impalas during opportunistic rather than one-on-one pursuit hunts<sup>37</sup>.

## Conclusions

The study shows that overall, the athletic capabilities of the two pursuit predators closely match their respective common prey, leading to a sustainable success rate, survival of both and reflecting an evolutionary arms race<sup>6,7</sup>. The predators have higher muscle power, are faster and have a greater capacity to accelerate and decelerate than their prey. The prey can match their predator's locomotor capabilities most closely through turning manoeuvrability, affording them a critical escape space. In evolutionary terms, there may be scope for further development of performance, for instance through increasing muscle power, but this specialization may be at the cost of locomotor economy, musculoskeletal robustness, or other ecologically relevant factors, such as prey capture ability, fighting or the capacity to adapt to a changing world.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 23 June 2017; accepted 2 January 2018.**

**Published online 24 January 2018.**

- Howland, H. C. Optimal strategies for predator avoidance: the relative importance of speed and manoeuvrability. *J. Theor. Biol.* **47**, 333–350 (1974).
- Moore, T. Y. & Biewener, A. A. Outrun or outmaneuver: Predator–prey interactions as a model system for integrating biomechanical studies in a broader ecological and evolutionary context. *Integr. Comp. Biol.* **55**, 1188–1197 (2015).
- Combes, S. A., Salcedo, M. K., Pandit, M. M. & Iwasaki, J. M. Capture success and efficiency of dragonflies pursuing different types of prey. *Integr. Comp. Biol.* **53**, 787–798 (2013).
- Domenici, P., Booth, D., Blagburn, J. M. & Bacon, J. P. Cockroaches keep predators guessing by using preferred escape trajectories. *Curr. Biol.* **18**, 1792–1796 (2008).
- Domenici, P., Blagburn, J. M. & Bacon, J. P. Animal escapology II: escape trajectory case studies. *J. Exp. Biol.* **214**, 2474–2494 (2011).
- Van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).
- Benton, M. J. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science* **323**, 728–732 (2009).
- Bro-Jørgensen, J. Evolution of sprint speed in African savannah herbivores in relation to predation. *Evolution* **67**, 3371–3376 (2013).
- Wilson, A. M. *et al.* Locomotion dynamics of hunting in wild cheetahs. *Nature* **498**, 185–189 (2013).
- Irschick, D. J. & Higham, T. E. *Animal Athletes: an Ecological and Evolutionary Approach* (Oxford Univ. Press, 2015).
- Broekhuis, F., Cozzi, G., Valeix, M., McNutt, J. W. & Macdonald, D. W. Risk avoidance in sympatric large carnivores: reactive or predictive? *J. Anim. Ecol.* **82**, 1098–1105 (2013).
- Schaller, G. B. *The Serengeti Lion: A Study of Predator–Prey Relations* (Univ. Chicago Press, 1972).
- Curtin, N. A., Woledge, R. C. & Aerts, P. Muscle directly meets the vast power demands in agile lizards. *Proc. R. Soc. Lond. B* **272**, 581–584 (2005).
- Kohn, T. A. & Noakes, T. D. Lion (*Panthera leo*) and caracal (*Caracal caracal*) type IIX single muscle fibre force and power exceed that of trained humans. *J. Exp. Biol.* **216**, 960–969 (2013).
- Williams, S. B., Tan, H., Usherwood, J. R. & Wilson, A. M. Pitch then power: limitations to acceleration in quadrupeds. *Biol. Lett.* **5**, 610–613 (2009).
- Carrier, D. R., Gregersen, C. S. & Silvertown, N. A. Dynamic gearing in running dogs. *J. Exp. Biol.* **201**, 3185–3195 (1998).
- Tan, H. & Wilson, A. M. Grip and limb force limits to turning performance in competition horses. *Proc. R. Soc. Lond. B* **278**, 2105–2111 (2011).

- Usherwood, J. R. & Wilson, A. M. Biomechanics: no force limit on greyhound sprint speed. *Nature* **438**, 753–754 (2005).
- Daley, M. A. in *Understanding Mammalian Locomotion: Concepts and Applications* Ch. 11 (ed. Bertram, J. E. A.) 277–306 (Wiley & Sons, 2016).
- Wilson, J. W. *et al.* Cheetahs, *Acinonyx jubatus*, balance turn capacity with pace when chasing prey. *Biol. Lett.* **9**, 20130620 (2013).
- Jindrich, D. L., Smith, N. C., Jespers, K. & Wilson, A. M. Mechanics of cutting maneuvers by ostriches (*Struthio camelus*). *J. Exp. Biol.* **210**, 1378–1390 (2007).
- Curtin, N. A., Diack, R. A., West, T. G., Wilson, A. M. & Woledge, R. C. Skinned fibres produce the same power and force as intact fibre bundles from muscle of wild rabbits. *J. Exp. Biol.* **218**, 2856–2863 (2015).
- Rome, L. C., Sosnicki, A. A. & Goble, D. O. Maximum velocity of shortening of three fibre types from horse soleus muscle: implications for scaling with body size. *J. Physiol. (Lond.)* **431**, 173–185 (1990).
- Seow, C. Y. & Ford, L. E. Shortening velocity and power output of skinned muscle fibers from mammals having a 25,000-fold range of body mass. *J. Gen. Physiol.* **97**, 541–560 (1991).
- Hill, A. V. The dimensions of animals and their muscular dynamics. *Science Progress* **38**, 209–230 (1950).
- West, T. G. *et al.* Power output of skinned skeletal muscle fibres from the cheetah (*Acinonyx jubatus*). *J. Exp. Biol.* **216**, 2974–2982 (2013).
- Crow, M. T. & Kushmerick, M. J. Chemical energetics of slow- and fast-twitch muscles of the mouse. *J. Gen. Physiol.* **79**, 147–166 (1982).
- Bartlam-Brooks, H. L., Bonyongo, M. C. & Harris, S. How landscape scale changes affect ecological processes in conservation areas: external factors influence land use by zebra (*Equus burchelli*) in the Okavango Delta. *Ecol. Evol.* **3**, 2795–2805 (2013).
- Schmidt-Nielsen, K. *Desert Animals. Physiological Problems of Heat and Water* (Clarendon, 1965).
- Wilson, A. M., McGuigan, M. P., Su, A. & van Den Bogert, A. J. Horses damp the spring in their step. *Nature* **414**, 895–899 (2001).
- McGuigan, M. P. & Wilson, A. M. The effect of gait and digital flexor muscle activation on limb compliance in the forelimb of the horse *Equus caballus*. *J. Exp. Biol.* **206**, 1325–1336 (2003).
- Pasi, B. M. & Carrier, D. R. Functional trade-offs in the limb muscles of dogs selected for running vs. fighting. *J. Evol. Biol.* **16**, 324–332 (2003).
- Carrier, D. R., Anders, C. & Schilling, N. The musculoskeletal system of humans is not tuned to maximize the economy of locomotion. *Proc. Natl Acad. Sci. USA* **108**, 18631–18636 (2011).
- Wynn, M. L., Clemente, C., Nasir, A. F. A. & Wilson, R. S. Running faster causes disaster: trade-offs between speed, manoeuvrability and motor control when running around corners in northern quolls (*Dasyurus hallucatus*). *J. Exp. Biol.* **218**, 433–439 (2015).
- Merz, A. The homicidal chauffeur. *AIAA J.* **12**, 259–260 (1974). <https://doi.org/10.2514/3.49215>
- Getz, W. & Pachter, M. Two-target pursuit–evasion differential games in the plane. *J. Optim. Theory Appl.* **34**, 383–403 (1981).
- Hubel, T. Y. *et al.* Energy cost and return for hunting in African wild dogs and cheetahs. *Nat. Commun.* **7**, 11034 (2016).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Amos for fabricating collars, N. Jordan and G. Giffillan, M. Claase and N. Terry and BPCT research assistants for working with us in the study area and M. Flyman (DWNP) for his support and enthusiasm; J. Usherwood, R. Bompfrey and A. R. Wilson for comments on the manuscript; EPSRC (EP/H013016/1), BBSRC (BB/J018007/1) and ERC (323041) for funding. The Botswana Predator Conservation Trust was supported by private donors, Tusk Trust and the Cincinnati Zoo. Work was approved by RVC Ethics & Welfare Committee (RVC 2013 1233) and Botswana Department of Wildlife and National Parks Research Permits were held by J.W.M. and A.M.W. (EWT 8/36/4 plus additions) and a Botswana Veterinary Registration held by A.M.W. Tissue shipping was covered by CITES, Botswana export, Botswana National Veterinary Laboratory approval, South African transit and UK DEFRA import permits.

**Author Contributions** A.M.W., T.Y.H. N.A.C., R.C.W. and T.G.W. conceived, designed and led the study. K.A.G., J.W.M., H.L.A.B.-B. and E.B. organized field work, monitored animals and downloaded data. A.M.W. performed veterinary procedures, J.C.L. and A.M.W. designed and built collars. R.D., M.L., N.A.C. and T.W. carried out muscle experiments and interpreted the muscle data. T.Y.H., O.P.D., T.G.W. and A.M.W. analysed data. S.W. created the model and carried out statistical analysis. A.M.W. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.M.W. ([awilson@rvc.ac.uk](mailto:awilson@rvc.ac.uk)).

**Reviewer Information** Nature thanks A. Biewener and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Animals.** All collared animals were located in northern Botswana with largely overlapping ranges (Fig. 1f). Animals were immobilized by free darting from a vehicle or helicopter mostly by A.M.W., using 80–100 mg ketamine and 2 mg medetomidine for cheetahs; 60 mg ketamine, 25 mg tiletamine hydrochloride, 25 mg zolazepam hydrochloride (as 50 mg zoletil, Virbac), 2 mg butorphanol tartrate and 6 mg medetomidine for lions; 1.5 mg thiafentanil oxalate, 2 mg butorphanol tartrate and 1,700 IU hyalase for impalas; and 7 mg etorphine hydrochloride, 50 mg azaperone and 1,700 IU hyalase for zebras. The reversal of immobilization of herbivores was done using diprenorphine or naltrexone at the end of the procedure and carnivores with atipamezole at 60 min after darting. While sedated, front and hind leg and body lengths were recorded. Collar data were downloaded by radio link every few weeks to a ground vehicle and collars were monitored. All animals were adult, nine lions (two male, seven female), five cheetahs (two male, three female), seven zebras (seven female), seven impalas (six male, one female). The lions and cheetahs were part of other ongoing projects in collaboration with the Botswana Predator Conservation Trust (<http://www.bpctrust.org>). Programmable drop-offs (two models, 108 g, Sirtrack Ltd or 50 g, Biotrack) were attached to the zebra and impala collars, respectively. Two drop-off units failed and collars were retrieved by re-darting.

Data were collected between April 2012 and November 2016 (cheetahs (June 2012–April 2013), lions (April 2012–June 2013), zebras (November 2014–September 2015), impalas (July–November 2016)). A subset of the data from cheetahs (367 out of 520 runs) were analysed in ref. 9.

**Muscle fibre measurements.** Biopsies were taken from the biceps femoris muscle by A.M.W., using a Bergstrom needle or conchotome forceps after collar placement. Animals were clipped, sterility was ensured and the biopsy site was treated with local antibiotics (200 mg cloxacillin, 75 mg ampicillin, Curaclox LC) and the animal was given analgesia (finadyne or metacam). Five male impala that had been killed for meat on a game ranch were dissected and provided additional muscle samples. Muscle samples were skinned by 30 min of immersion in ice-cold relaxing solution containing 2% Triton X-100 and exposed to a pH-6 relaxing solution to inactivate any foot-and-mouth disease virus. Triton X-100 was washed out with fresh relaxing solution and samples were immersed in 500 mM trehalose containing 0.5% glycerol<sup>22</sup>, frozen in liquid nitrogen and stored in an IATA-approved dry-shipper (Biotrek 3 Statebourne Cryogenics) for transport to the United Kingdom. In the United Kingdom, biopsies were stored at  $-80^{\circ}\text{C}$ . Periodically, individual biopsies were thawed and had cryopreserving trehalose replaced with a relaxing solution. Our previous work showed that biopsies stored for 20 months using this protocol showed no discernible loss of mechanical power<sup>22</sup>. Thawed biopsies were stored at  $-20^{\circ}\text{C}$  in a relaxing solution made up in glycerol and used for fibre preparation and testing within four weeks.

Fibre fragments were first suspended while in relaxing solution between the motor and force transducer of a 600A permeabilized-fibre apparatus (Aurora Scientific). T-shaped aluminium clips were attached to fibre ends and used to suspend fibres from steel wire hooks that were glued with shellac to the motor and transducer. Fibres were visualized using a 900B digital camera (Aurora Scientific). The camera image was used to set the sarcomere length (SL) of a fibre fragment to between 2.5 and 2.6  $\mu\text{m}$ . Fibre length ( $L_0$ ), depth and width were then measured in mm. The fibre cross-sectional area was calculated for each fibre, assuming an elliptical shape.

Single skinned fibres were activated by temperature jump, from  $1^{\circ}\text{C}$  to  $25^{\circ}\text{C}$  (Extended Data Fig. 2a), using approaches similar to those previously described<sup>22</sup>. The composition and ionic strength (200 mM) of the various solutions was as previously described<sup>22</sup>. To activate a fibre, it was immersed consecutively in solutions for low-temperature pre-activation (for 45 s), low-temperature activation (for 4 s), high-temperature activation (6 s) and high-temperature relaxation. The example in Extended Data Fig. 2a shows the time courses of solution changes and force responses for an impala fibre, starting from the final 3 s of cold-temperature pre-activation. The force baseline at  $0.7 L_0$  was recorded in high-temperature relaxing solution before re-setting  $L_0$  to the original starting length and checking SL.

The standard procedure to measure fibre power and determine peak power was modified to perform four different force-control events during each 6-s activation to deliver more data per fibre<sup>22</sup>. In brief, after temperature jump to  $25^{\circ}\text{C}$ , force developed to a plateau at constant length (isometric force), and the fibre was then clamped for 20 ms to a predetermined fraction of peak isometric force—the actual force achieved in the first force clamp was calculated by the 600A based on the difference between baseline force (measured and stored within a 600A protocol just before  $1^{\circ}\text{C}$  activation) and isometric force (measured and stored within a 600A protocol just before onset of a force clamp). The shortening velocity during

force clamp was measured and used to calculate power output. The fibre was then released to slack length in order to re-measure the baseline force and it was then lengthened to  $L_0$  over a period of 5 ms. This step avoids a high eccentric force transient during lengthening. The baseline measurement was saved again in the 600A protocol and used to compare with the measurement of stable isometric force achieved after lengthening the fibre to  $L_0$ —the force attained in the next force clamp was based on the difference between the newly saved values of baseline and isometric force. Four different force-control events were conducted during the 6-s activation at  $25^{\circ}\text{C}$  (see Extended Data Fig. 2a). Examples of a force clamp and of the fibre-length changes required to hold force constant are shown in Extended Data Fig. 2b, c. Relaxation and a final force baseline-check were also conducted at  $25^{\circ}\text{C}$  (Extended Data Fig. 2a). Three activations provided up to twelve different force-clamp measurements in order to quantify a power versus force relationship and peak power for each fibre (see below).

A fibre was counted as 'tested' if it did not break on the test apparatus, and if test conditions (solution temperature and chemistry) were maintained as prescribed in the experimental design. A fibre was counted as 'included' in the mechanical tests if the maximum isometric force  $>75$  kPa, and during the repeated activations, isometric force for each test remained  $>80\%$  of the peak isometric force observed. For each fibre, we conducted three activations, each with four force-control, or shortening, events in order to collect, at most, 12 points for a power-force curve fit. An individual data point (that is, a single force-control event) could be rejected (either because of poor (for example, unstable/oscillating) force during fibre-shortening or because of low ( $<80\%$  of maximum) isometric force), but exclusion of a data point on this basis will not necessarily have caused the fibre to have been rejected, unless the spread of usable data points was insufficient for curve fitting. Of the 209 fibres initially tested with apparent success, 35 were excluded (three out of 40 cheetah fibres, 14 out of 64 lion fibres, 14 out of 71 zebra fibres, four out of 34 impala fibres).

The data for each fibre were analysed as described previously<sup>22</sup>.

$$Q = \frac{F_{\text{max}}^2 Q_{\text{max}}^2 \left(1 - \frac{F_0}{F_{\text{max}}}\right)}{1 + F_0(F_0 - 2F_{\text{max}})/F_{\text{max}}^2} \quad (1)$$

Equation (1) describes the dependence of relative power ( $Q = \text{power} / (F_{\text{isom}} \times L_0)$ ; in  $\text{s}^{-1}$ ) on relative force (force during shortening /  $F_{\text{isom}}$ ; units are dimensionless). Peak relative power ( $Q_{\text{max}}$ ) was found by fitting a line to the data by adjusting three parameters:  $Q_{\text{max}}$ , the force intercept ( $F_0$ ) and force at peak power ( $F_{\text{max}}$ ). An example plot and best-fit curve is shown in Extended Data Fig. 2d. Peak power in  $\text{W kg}^{-1}$  is obtained by multiplying  $Q_{\text{max}}$  by maximum isometric stress in kPa and dividing by fibre density,  $1.064 \text{ g ml}^{-1}$  (ref. 38).

After mechanical tests, the 'low-performing' single skinned fibres were pinned onto a gelatine base in cryomolds, flooded with OCT (Tissue-Tek) and frozen in liquid nitrogen. Sections ( $8\text{-}\mu\text{m}$  thick) were cut and immunostained with mouse anti-MHC fast monoclonal antibody (MY-32, 1:1,000; ab51263, Abcam) for type-II fibres, and mouse anti-MHC slow monoclonal antibody (1:50; MAB1628, Merck Millipore) for type-I fibres.

**Muscle data statistics.** A linear mixed-effects model was fitted in R (R Foundation for Statistical Computing) for peak power, velocity at peak power, stress at peak power and isometric stress against a factor distinguishing predator and prey with the interaction of this factor with a categorical variable 'performance classification'<sup>39</sup>. Within the factor distinguishing predator and prey, we included a nested random effect by subject and fibre. The residuals of this model exhibited heteroscedasticity and so the variance of the error term was allowed to vary by performance classification. General linear hypothesis tests were then performed. **Temperature.** Muscle power is highly temperature-dependent<sup>25</sup> and in a previous study<sup>26</sup>, using their data and literature data, it was shown that for the temperature range of  $20\text{--}35^{\circ}\text{C}$ , a temperature coefficient ( $Q_{10}$ ; ratiometric increase in rate with a temperature increase of  $10^{\circ}\text{C}$ ) of 2.3 is appropriate<sup>40,41</sup> (see figure 7 of ref. 26). A  $Q_{10}$  of 2.3 was used to predict powers at a body temperature of  $38^{\circ}\text{C}$ .

**Collar design.** All collars were designed and constructed in-house and are described in detail in ref. 9 and in the Supplementary Information.

Ethical guidelines suggest a collar mass limit of 5–10% of the body mass<sup>42</sup> to minimize the effect on the animal; our collars were below that threshold at 0.3 to 1.0% (collar mass: cheetah, 340 g; lion, 970 g; zebra, 930 g and impala, 450 g. Drop-off mechanisms were 108 g (Sirtrack) and 50 g (Biotrack)). The electronics package was similar in all collar versions with almost identical functionality.

**Signal processing.** GPS-INS processing was used to reduce noise and improve precision for the position and velocity analysis, as well as increasing the temporal resolution of the data. GPS and IMU measurements were fused<sup>43</sup> using a 12-state extended Kalman filter<sup>43</sup> in loosely coupled architecture written in MATLAB (The Mathworks). The total state formulation used propagates position, velocity and

orientation states with time using the IMU measurements in a simplified form of the strap-down inertial navigation equations<sup>44</sup>. The associated process noise was estimated from the known error characteristics of the inertial sensors used. GPS position and velocity updates were used as measurement updates, and receiver accuracy data for each fix used to estimate measurement noise to appropriately weight the GPS to the inertial solution.

The filter was run in reverse time from the last GPS observation of each run to the beginning of the buffered inertial data. During the short time period in which only inertial data was present (delay between trigger and first GPS fix), the filter propagation was equivalent to open-loop inertial navigation. The filter was initialized using the last GPS position and velocity data, and Euler angles assumed zero with covariances appropriate for the uncertainty in that assumption. A Rauch–Tung–Striebel smoother<sup>45</sup> was then applied in forward time on the Kalman-filtered data. This is equivalent to combining backward and forward solutions, effectively halving the open-loop INS integration period between GPS observations. In cases for which it was not possible to reconstruct the period before the first GPS observation (time too long or GPS accuracy insufficient), runs start at medium speeds rather than very low speeds.

**Calculation of speed and stride times.** Vertical accelerations were used to determine stride times. A zero phase band pass Butterworth filter (fourth order) was applied with cut-off frequencies of 1 Hz and 6.6 Hz (twice the maximum stride frequency in the cheetah and impala). A peak detection function was used to detect peaks with a minimum period of 0.25 s between peaks and a minimum peak height of 0.1 g.

Species-specific gait parameters, such as transition speeds and expected stride frequencies for walking and trotting (based on ref. 46), were used to remove double peaks in strides in symmetrical gaits. Horizontal stride speed was derived from the Kalman-filtered velocity averaged over strides in order to remove the effects of speed fluctuation through the stride and collar oscillation relative to the centre of mass.

**Tangential acceleration, change of heading and centripetal acceleration over stride.** Stride times were used to calculate tangential (fore–aft) acceleration, centripetal (turning) acceleration and change in heading between strides. The displacement vectors between consecutive strides were then calculated:

$$\mathbf{U}_i = \mathbf{P}_i - \mathbf{P}_{i-1} \quad (2)$$

where  $\mathbf{P}_i$  is the two-dimensional position at stride  $i$ .

Change of heading ( $\Delta\theta_i$ ) was calculated from the angle between the two vectors:

$$\Delta\theta_i = \sin^{-1} \left( \frac{|\mathbf{U}_{i+1} \times \mathbf{U}_i|}{|\mathbf{U}_{i+1}| |\mathbf{U}_i|} \right) \quad (3)$$

Angular velocity ( $\omega_i$ ) was derived by dividing the change of heading by the time between mid-stride positions  $\Delta T$ :

$$\omega_i = \frac{\Delta\theta_i}{\Delta T} \quad (4)$$

The tangential or fore–aft acceleration ( $a_{t,i}$ ) and centripetal acceleration ( $a_{c,i}$ ) were then computed from mid-stride speeds  $v_i$ :

$$a_{t,i} = \frac{v_{i+1} - v_i}{\Delta T} \quad (5)$$

$$a_{c,i} = \frac{v_i^2}{r_i} = \omega_i v_i \quad (6)$$

Negative values for tangential acceleration represent deceleration. Absolute values were used for centripetal acceleration, equalling right and left turns. For visual purposes the data in Fig. 3n were mirrored around the vertical axis.

Mass-specific centre of mass (COM) stride work (net COM kinetic energy change in a stride) was calculated as change in speed over a stride multiplied by stride average speed. Mass-specific COM power was calculated as the dot product of stride-averaged tangential acceleration and stride-averaged velocity (that is, multiply forward acceleration by forward speed):

$$P_{t,i} = a_{t,i} v_i \quad (7)$$

**Calculation of grip limits.** Grip limits are shown in Fig. 3. Friction poses a limit on acceleration and is the product of friction coefficient  $\mu$  and force normal to the surface (based on acceleration due to gravity,  $g$ ). Therefore the maximum total horizontal acceleration  $a_{\max}$  is limited to:

$$a_{\max} = \mu g \quad (8)$$

where  $a_{\max}$  is the resultant (combination) of tangential and centripetal acceleration:

$$\begin{aligned} a_{c,\max} &= \mu g \\ a_{t,\max} &= \mu g \end{aligned} \quad (9)$$

Maximum turning speed  $v_{\max}$  depends on friction, gravity and turning radius and is calculated based on equations (6) and (9):

$$v_{\max} = \sqrt{\mu g r} \quad (10)$$

**Calculation of stride frequency.** Regression lines were fitted to stride frequency versus speed data at running speeds. Sections with running data were identified using an unsupervised clustering algorithm on three features derived from windows of accelerometer signals (4-s long)<sup>46</sup>. Features were chosen on the basis of domain knowledge and were the s.d. of the horizontal and vertical axis accelerometer signals and an autocorrelation estimate of the stride frequency<sup>46</sup>. Features were normalized to have zero mean and unit standard deviation before they were clustered using the  $k$ -means algorithm. The number of clusters was determined using the Davis–Bouldin criterion<sup>47</sup> and human inspection. Subsequently, the sections identified to contain running data were isolated and vertical acceleration was used to determine stride times (see ‘Calculation of speed and stride times’), stride frequency was calculated from the time between acceleration peaks. Regression lines were calculated for the subgroup from each bin representing the middle 60%, the highest 20% of positive and highest 20% of negative power (Fig. 3b, g).

**Maximum performance analysis.** Extracting values that reflect maximum performance carries the risk of choosing outliers generated by non-Gaussian GPS noise rather than realistic values. Previous work reduced the risk of overestimating performance by weighting stride parameters, such as stride speed and accelerations, by the previous and following stride<sup>9,37,48</sup>. Here we chose a different approach, not weighting, but calculating the 98th percentile for each of a number of bins (Fig. 3) in order to also address the effect of different sample sizes and accelerations that were not sustained for three consecutive strides. In addition, obvious errors (speeds  $>30 \text{ m s}^{-1}$  and total stride averaged accelerations exceeding a magnitude of  $20 \text{ m s}^{-2}$ ) were removed from the dataset.

An inherent issue with comparing the performance of different species lies in their different movement patterns, with lion and zebra having a considerably higher proportion of straight, constant low speed strides than impala and cheetah. In order to extract manoeuvring strides, a cut-off based on the magnitude of the total horizontal acceleration (combined tangential and centripetal acceleration) was performed. This cut-off could not be universal, because different animals had different amounts of low-speed steady-state behaviour in their accelerometer traces. This manifested itself in large differences between species in kurtosis of the acceleration-distribution histograms. To address these differences a species-specific cut-off was used. To ensure that this cut-off still gave comparable results for the different species, the characteristic scale of the kurtosis for each distribution was estimated using:

$$s_\alpha = \sigma_\alpha k_\alpha^{1/4} \quad (11)$$

where  $s_\alpha$  is our characteristic scale,  $\sigma_\alpha$  is the standard deviation and  $k_\alpha$  is the Pearson’s kurtosis, all for species  $\alpha$ . If a cut-off of  $c_\alpha$  was used for one species, then we can calculate the cut-off for species  $\beta$  by:

$$c_\beta = \frac{s_\beta}{s_\alpha} c_\alpha \quad (12)$$

The effect of this cut-off on the distribution of total horizontal acceleration is shown in Extended Data Fig. 3a, b.

In Fig. 3n, tangential acceleration is plotted against centripetal accelerations. The Cartesian coordinates were transformed into polar coordinates in order to bin the data. Calculations were performed on absolute centripetal acceleration values to boost data point numbers in bins and then mirrored on the vertical axis; the semicircle was divided into a total of six bins.

The cut-off was adjusted, so that the number of data points in a bin was at least 200 for all species. The cut-off was determined by the impala, which had the lowest number of data points.

The parameters were plotted versus horizontal speed (except for stride frequency) and the 98th percentile was calculated for each of a number of speed bins for which the width was defined by the requirement that each bin should include 400 data points. The final (highest speed) bins with less than 400 data points were discounted. In Fig. 3 a moving average of three bins was applied to all data except Fig. 3n. Data were interpolated to allow the calculation of species performance ratios (Fig. 3) at  $1 \text{ m s}^{-1}$  speed positions.

**Statistical analysis.** The maximum performance of the predator and prey were compared using a set of linear models of maximum positive and negative power, positive and negative tangential acceleration and absolute centripetal acceleration.

The maximum performance of each individual was quantified by taking the 98th percentile of the positive and negative tangential acceleration and absolute centripetal acceleration of each individual.

Negative and positive power covaried with speed and was binned by speed as above and the 98th percentile within each bin was computed for each subject within a species. A linear regression was then performed and the predicted power at  $8 \text{ m s}^{-1}$  calculated for each subject.

Linear models were fitted to these data using restricted maximum likelihood, with the maximum powers and accelerations as dependent variables, against a factor for predator versus prey and a factor for each pairing (zero for cheetah–impala and one for zebra–lion). Models with an interaction term between these two factors were fitted, but comparing these models to those previously described (this time fitted using maximum likelihood) indicated the interaction term was superfluous (effect sizes were small and associated  $P$  values not significant). The interaction term was therefore dropped from the analysis in all models, except when comparing stride frequencies for which there was a substantial interaction term (effect sizes large and associated  $P$  values significant). Here, the model was fitted by individual species pairs. Owing to the presence of heteroscedasticity, the error term was allowed to vary for each species.

**Chase–evasion model.** The model combines the observed acceleration capacity with a maximum speed constraint to produce possible position profiles for predators and prey in the subsequent two strides of a chase. We simulate the possible positions of the prey given the prey's initial speed. We then do the same for the predator, optimizing the predator's initial speed to give maximum overlap in final positions of the predator and prey.

We begin with the observed maximum accelerations for our subjects (Extended Data Table 2b). We approximate the possible impulsive accelerations of the animals by assuming they have a maximum tangential acceleration forward,  $a_t$ , a maximum reverse tangential acceleration,  $a_{tr}$ , and a maximum centripetal acceleration,  $a_c$ . The profile of possible accelerations is assumed to be two half-ellipses with the semi-minor axis along the direction of motion. The top ellipse has semi-minor axis radius  $a_t$ , the bottom ellipse has semi-minor axis radius  $a_{tr}$ , and both have a semi-major axis of length  $a_c$ .

The animals are assumed to have a maximum possible speed,  $v$ . This places a further constraint on the possible profile of accelerations as no acceleration can result in a speed above this maximum.

To find this constraint, we assume that a predator and its prey are galloping at a common stride frequency (Fig. 3b, g) and phase, and that the bulk of the impulse that they can achieve in a stride is performed in a short duration (stance). On any given stride the animal can apply an impulse to change direction, subject to the constraint that the resulting speed cannot be greater than the animal's maximum speed,  $v_M$ . If the animal is at a speed  $v$  along a unit direction  $\hat{i}$  and an impulsive acceleration  $a_0\hat{i} + a_1\hat{j}$  with  $\hat{j}$  perpendicular to  $\hat{i}$  is to be applied then the resulting speed is:

$$\left(v + \frac{a_0}{f}\right)^2 + \left(\frac{a_1}{f}\right)^2 \quad (14)$$

where  $f$  is the stride frequency. This must be less than  $v_M$ . This implies a pair of quadratic relations between  $a_0$  and  $a_1$  subject to  $v$  and  $v_M$  of the form:

$$a_0 = \pm \sqrt{(v_M^2 - v^2)f^2 - a_1^2 - 2a_1vf} \quad (15)$$

The simulation allows our subjects to accelerate to anywhere within the area formed by the union of the area above the negative root of this equation, below the positive root, and within the two half ellipses previously mentioned.

We note that the possible acceleration profile depends both on the position of the animal and its current speed (an animal that is slow will not be constrained by its maximum possible speed, whereas one going at its maximum speed cannot accelerate forward). This means that simulating the animal's possible positions forward in time increases in complexity with each stride taken, as both the new position of the animal and the new speed must be retained. As such we confined ourselves to simulating two strides forward from our starting conditions; that is, we are only concerned with strategies for the predator and prey at the very end of a chase.

We assume that the prey performs an evasive acceleration on the first stride, while the predator continues to chase without changing velocity. On the second stride, the prey again accelerates, and now the predator also has the ability to react to the acceleration it observes in the first stride. We ran 100 such simulations for starting separations varying from the maximal separation that makes

capture possible within two strides down to half a predator length separation (cheetah =  $0.66 \text{ m}^{49}$ , lion =  $0.92 \text{ m}^{50}$ ). If the prey and predator are closer than this, then the predator is already close enough for prey capture.

For a given prey speed and initial predator–prey separation, we find that predator speed, which maximizes the capture probability by means of a Nelder–Mead simplex optimization, is subject to the constraint that the initial predator speed must be greater than or equal to the initial prey speeds. Owing to ambiguity in the solution space, a small penalty term encouraging faster speeds from the predator was added in the form  $\text{eps} \times v_{\text{pred}}$ , with  $\text{eps} = 10^{-6}$ . This ensured that for data with a range of optimal best speeds for the predator, the fastest was selected. This had no effect on the value of the optima up to four significant digits.

To test how a change in predator or prey performance influences hunt outcome, we adjusted the performance by multiplying the maximum recorded tangential and the centripetal accelerations of the prey or predator by a number ranging from 0.6 to 1.4 to deliver values to insert into the model for animals of greater or lesser athleticism, respectively. This number is the  $x$ -axis performance adjustment in Fig. 4g, h and rerunning the simulation to obtain capture probabilities. Maximum speed was not adjusted.

**List of symbols.** *Muscle studies.*  $F_{\text{isom}}$ , fibre isometric force;  $F$ , fibre force during shortening;  $SL$ , fibre sarcomere length;  $L_0$ , fibre length when the sarcomere length is set to  $2.55 \mu\text{m}$ ;  $Q$ , fibre relative power;  $Q_{\text{max}}$ , fitted maximum relative power;  $F_0$ , fitted force intercept on a distribution of  $Q$  against  $F/F_{\text{isom}}$ ;  $FQ_{\text{max}}$ , fitted relative force at maximum power;  $CSA$ , fibre cross-sectional area;  $Q_{10}$ , the ratiometric increase in rate with a  $10^\circ\text{C}$  temperature change.

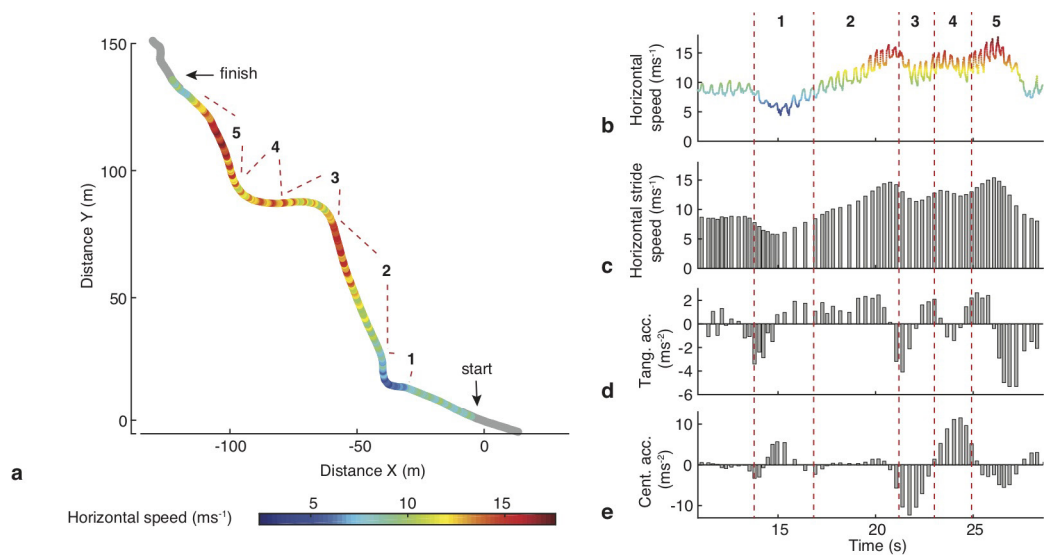
*Locomotion and model.*  $i$ , stride number;  $P_i$ , two-dimensional position;  $U_i$ , two-dimensional position difference between subsequent strides;  $\Delta\theta_i$ , signed change of heading;  $\omega_i$ , heading angular velocity;  $\Delta T$ , sampling interval;  $a$ , total horizontal acceleration;  $a_t$ , tangential or forward acceleration,  $a_{tr}$ , tangential reverse acceleration;  $a_c$ , centripetal acceleration,  $a_0$  and  $a_1$  are generic accelerations;  $r$ , turn radius;  $v$ , stride-averaged horizontal speed,  $v_{\text{max}}$ , maximal turning speed,  $v_M$ , maximum speed;  $P_p$ , mass-specific fore–aft power;  $\mu$ , coefficient of friction;  $m$ , body mass;  $g$ , gravity;  $\alpha$  and  $\beta$ , species indices;  $s$ , characteristic scale;  $\sigma$ , s.d.;  $k$  is the Pearson's kurtosis.

**Code availability.** Python code for the simulation predator–prey model and data are available as Supplementary Data files.

**Data availability.** The authors declare that all relevant individual fibre and individual stride (processed, not raw) data supporting the findings of this study are available as Supplementary Data or as Source Data. Any further data are available from the corresponding author upon reasonable request.

38. Méndez, J. & Keys, A. Density and composition of mammalian muscle. *Metabolism* **9**, 184–188 (1960).
39. Pinheiro, J. & Bates, D. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000).
40. Ranatunga, K. W. Temperature dependence of mechanical power output in mammalian (rat) skeletal muscle. *Exp. Physiol.* **83**, 371–376 (1998).
41. Barclay, C. J., Woledge, R. C. & Curtin, N. A. Is the efficiency of mammalian (mouse) skeletal muscle temperature dependent? *J. Physiol. (Lond.)* **588**, 3819–3831 (2010).
42. Sikes, R. S. et al. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J. Mamm.* **97**, 663–688 (2016).
43. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
44. Titterton, D. & Weston, J. L. *Strapdown Inertial Navigation Technology*. 2 edn, Vol. 207 (AIAA, 2004).
45. Rauch, H. E., Striebel, C. & Tung, F. Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3**, 1445–1450 (1965).
46. Dewhurst, O. P. et al. An exploratory clustering approach for extracting stride parameters from tracking collars on free ranging wild animals. *J. Exp. Biol.* **220**, 341–346 (2017).
47. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-1**, 224–227 (1979).
48. Hubel, T. Y. et al. Additive opportunistic capture explains group hunting benefits in African wild dogs. *Nat. Commun.* **7**, 11033 (2016).
49. Estes, R. D. *The Behaviour Guide to African Mammals: including hoofed Mammals, Carnivores, Primates* (Russell Friedman Books, 1991).
50. Nowak, R. M. *Walker's Mammals of the World* Vol. 1 (JHU, 1999).
51. Williams, S. B., Wilson, A. M., Rhodes, L., Andrews, J. & Payne, R. C. Functional anatomy and muscle moment arms of the pelvic limb of an elite sprinting athlete: the racing greyhound (*Canis familiaris*). *J. Anat.* **213**, 361–372 (2008).
52. Williams, S. B., Wilson, A. M., Daynes, J., Peckham, K. & Payne, R. C. Functional anatomy and muscle moment arms of the thoracic limb of an elite sprinting athlete: the racing greyhound (*Canis familiaris*). *J. Anat.* **213**, 373–382 (2008).
53. Webster, E. L., Hudson, P. E. & Channon, S. B. Comparative functional anatomy of the epaxial musculature of dogs (*Canis familiaris*) bred for sprinting vs. fighting. *J. Anat.* **225**, 317–327 (2014).
54. Hudson, P. E. et al. Functional anatomy of the cheetah (*Acinonyx jubatus*) forelimb. *J. Anat.* **218**, 375–385 (2011).

55. Hudson, P. E. *et al.* Functional anatomy of the cheetah (*Acinonyx jubatus*) hindlimb. *J. Anat.* **218**, 363–374 (2011).
56. Payne, R. C., Hutchinson, J. R., Robilliard, J. J., Smith, N. C. & Wilson, A. M. Functional specialisation of pelvic limb anatomy in horses (*Equus caballus*). *J. Anat.* **206**, 557–574 (2005).
57. Payne, R. C., Veenman, P. & Wilson, A. M. The role of the extrinsic thoracic limb muscles in equine locomotion. *J. Anat.* **205**, 479–490 (2004).
58. Crook, T. C. *et al.* Comparative anatomy and muscle architecture of selected hind limb muscles in the quarter horse and Arab. *J. Anat.* **212**, 144–152 (2008).
59. Williams, S. B., Payne, R. C. & Wilson, A. M. Functional specialisation of the pelvic limb of the hare (*Lepus europeus*). *J. Anat.* **210**, 472–490 (2007).
60. Williams, S. B., Wilson, A. M. & Payne, R. C. Functional specialisation of the thoracic limb of the hare (*Lepus europeus*). *J. Anat.* **210**, 491–505 (2007).
61. Smith, N. C., Wilson, A. M., Jespers, K. J. & Payne, R. C. Muscle architecture and functional anatomy of the pelvic limb of the ostrich (*Struthio camelus*). *J. Anat.* **209**, 765–779 (2006).
62. Schaller, N. U. *Structural Attributes contributing to Locomotor Performance in the Ostrich*. PhD thesis, Univ. Heidelberg (2008).
63. Cuff, A. R. *et al.* The scaling of postcranial muscles in cats (Felidae) II: hindlimb and lumbosacral muscles. *J. Anat.* **229**, 142–152 (2016).
64. Cuff, A. R. *et al.* The scaling of postcranial muscles in cats (Felidae) I: forelimb, cervical, and thoracic muscles. *J. Anat.* **229**, 128–141 (2016).
65. Holland, R., Loveday, D. & Ferguson, K. How much meat to expect from a beef carcass. <http://extension.tennessee.edu/publications/Documents/PB1822.pdf> (Univ. Tennessee, Institute of Agriculture, 2014).
66. Sacks, R. D. & Roy, R. R. Architecture of the hind limb muscles of cats: functional significance. *J. Morphol.* **173**, 185–195 (1982).
67. Shahar, R. & Milgram, J. Morphometric and anatomic study of the hind limb of a dog. *Am. J. Vet. Res.* **62**, 928–933 (2001).
68. Gunn, H. M. Muscle, bone and fat proportions and muscle distribution of thoroughbreds and other horses. In *Proc. 2nd ICEEP Equine Exercise Physiology* (Gillespie, J. R. & Robinson, N. E.) 253–264 (ICEEP, 1988).
69. Alexander, R. M. *Principles of Animal Locomotion* (Princeton Univ. Press, 2003).
70. Lindstedt, S. L., LaStayo, P. C. & Reich, T. E. When active muscles lengthen: properties and consequences of eccentric contractions. *News Physiol. Sci.* **16**, 256–261 (2001).
71. Roberts, T. J. & Azizi, E. The series-elastic shock absorber: tendons attenuate muscle power during eccentric actions. *J. Appl. Physiol.* **109**, 396–404 (2010).
72. Abbott, B. C., Bigland, B. & Ritchie, J. M. The physiological cost of negative work. *J. Physiol. (Lond.)* **117**, 380–390 (1952).

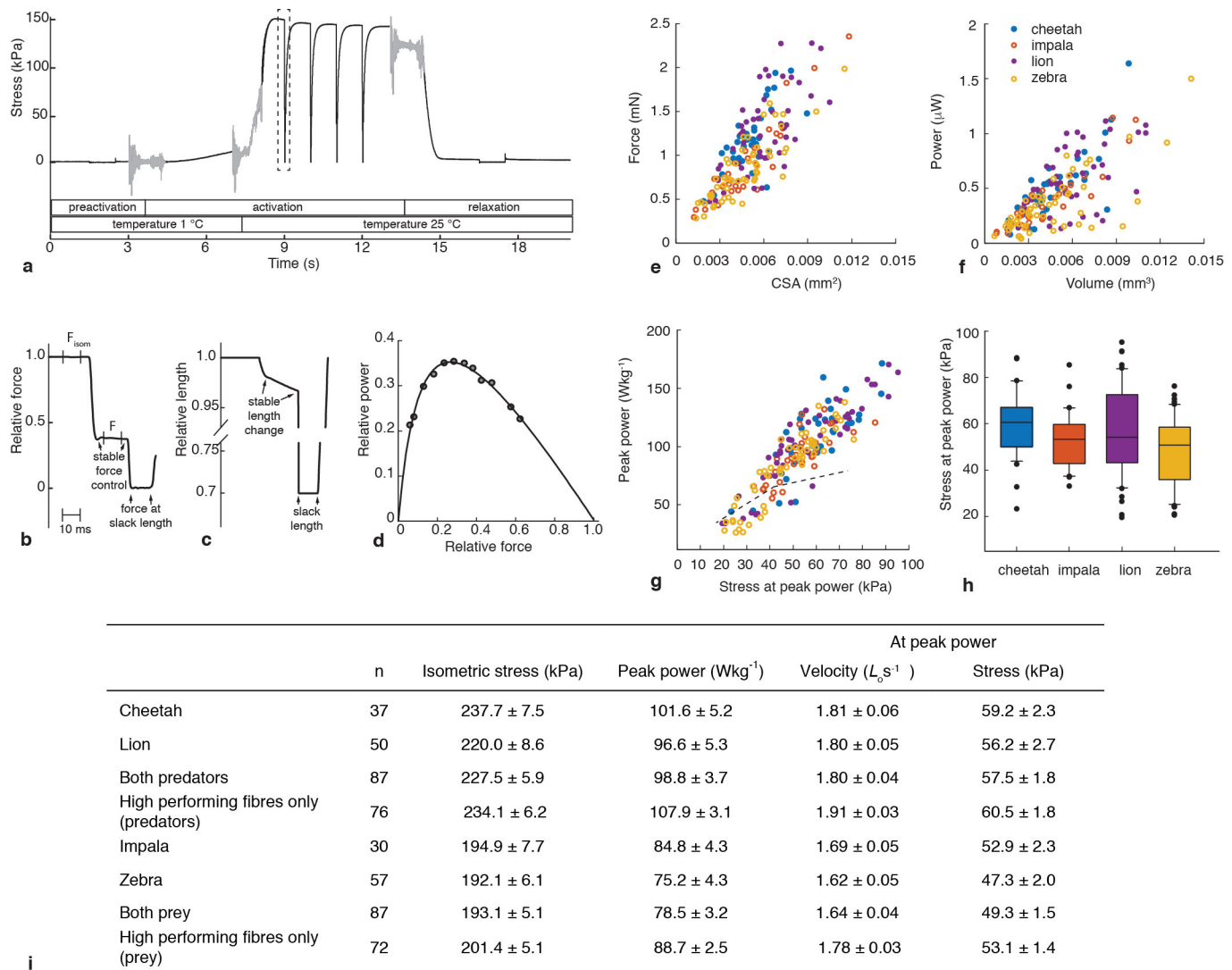


| Species | Number of individuals | Number of trials | Total strides | Strides used (non-steady state) | Number of strides per run (mean±s.d.) | Distance per run (m) (mean±s.d.) |
|---------|-----------------------|------------------|---------------|---------------------------------|---------------------------------------|----------------------------------|
| Cheetah | 5                     | 520              | 23871         | 7509                            | 50 ± 25                               | 157 ± 122                        |
| Impala  | 7                     | 515              | 22491         | 8884                            | 47 ± 33                               | 151 ± 149                        |
| Lion    | 9                     | 2726             | 101110        | 15947                           | 39 ± 22                               | 80 ± 76                          |
| Zebra   | 7                     | 1801             | 64952         | 14089                           | 38 ± 39                               | 69 ± 91                          |

f

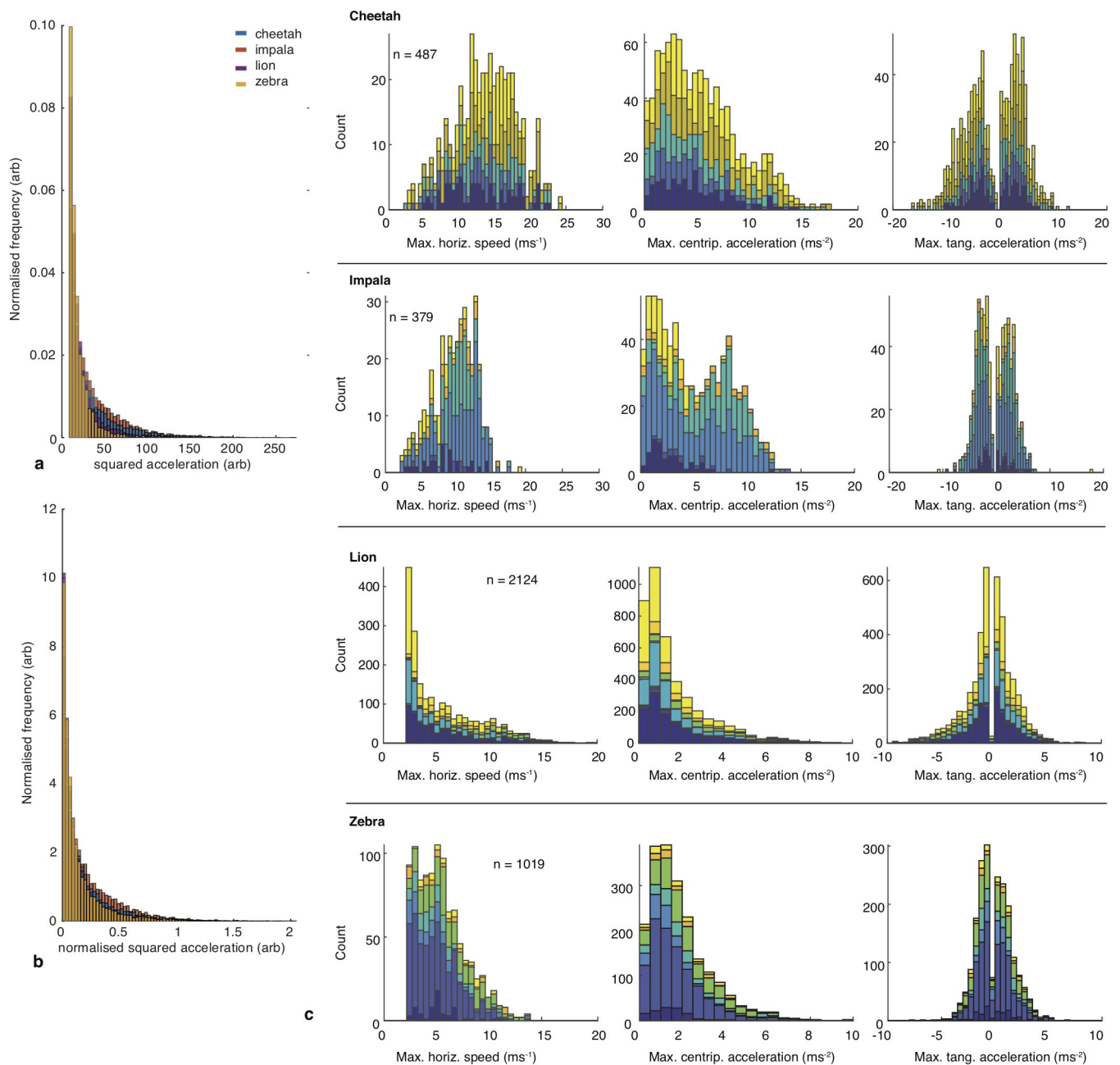
**Extended Data Figure 1 | GPS data summary.** **a**, Example manoeuvring sequence for a cheetah showing position based on fused GPS-IMU data (250 Hz) colour-coded according to speed and segmented for clarity (1–5, duplicated in **b–e**). **b**, Speed based on fused GPS-IMU data (250 Hz). **c–e**, Stride-wise values for speed (averaged over stride), tangential (fore–aft) acceleration (change in stride speed/stride durations) and centripetal

(lateral) acceleration ((change in heading/stride duration) × stride speed). **f**, Details on the animals used and datasets collected. Reduced dataset of non-steady state stride used for analysis of maximum performance. Note, the number of strides and distance per run was based on all strides (steady-state included).



**Extended Data Figure 2 | Muscle data summary.** **a–c**, Time course of stress (force) development in a single skinned impala fibre, showing transition of the fibre through pre-activation, activation and relaxation solutions, and stress development after temperature jump (T-jump) from 1 to 25°C. The sample rate was 5 kHz. The grey noisy parts of the stress trace denote periods of solution change. The four downward 'spikes' in the stress record (at 9, 10, 11 and 12 s) are distinct periods of force control, where the fibre length was first rapidly reduced from  $L_0$  and then reduced at an appropriate rate to maintain force at pre-defined sub-maximal levels. The broken-line box in **a** surrounds the first episode of force control and is presented in **b** and **c** on an expanded time scale. **b** Relative force ( $F/F_{\text{isom}}$ ) was reduced to 40% of the maximum for 20 ms, where  $F$  indicates force during shortening,  $F_{\text{isom}}$  indicates isometric force. Isometric force ( $F_{\text{isom}}$ ) and the force during force clamp ( $F$ ) were recorded as average values for the central 10-ms intervals (vertical lines). A force measurement,  $F_{\text{isom}}$ , was recorded just before each of the four force-control events and used in the calculation of  $F/F_{\text{isom}}$ . **c**, Shortening speed (in  $L_0 \text{ s}^{-1}$ ) was derived from the rate of change in fibre length during each force clamp. At the end of the force clamp, fibre length was 'quick-released' to a slack length (70% of  $L_0$ ), where force was reduced to the zero baseline. After 10 ms at slack length, the motor lengthened the fibre back to the starting length ( $L_0$ ), isometric stress was re-established (as shown in **a**) and another force control event was initiated. **d**, Twelve points on a power-force relationship could be

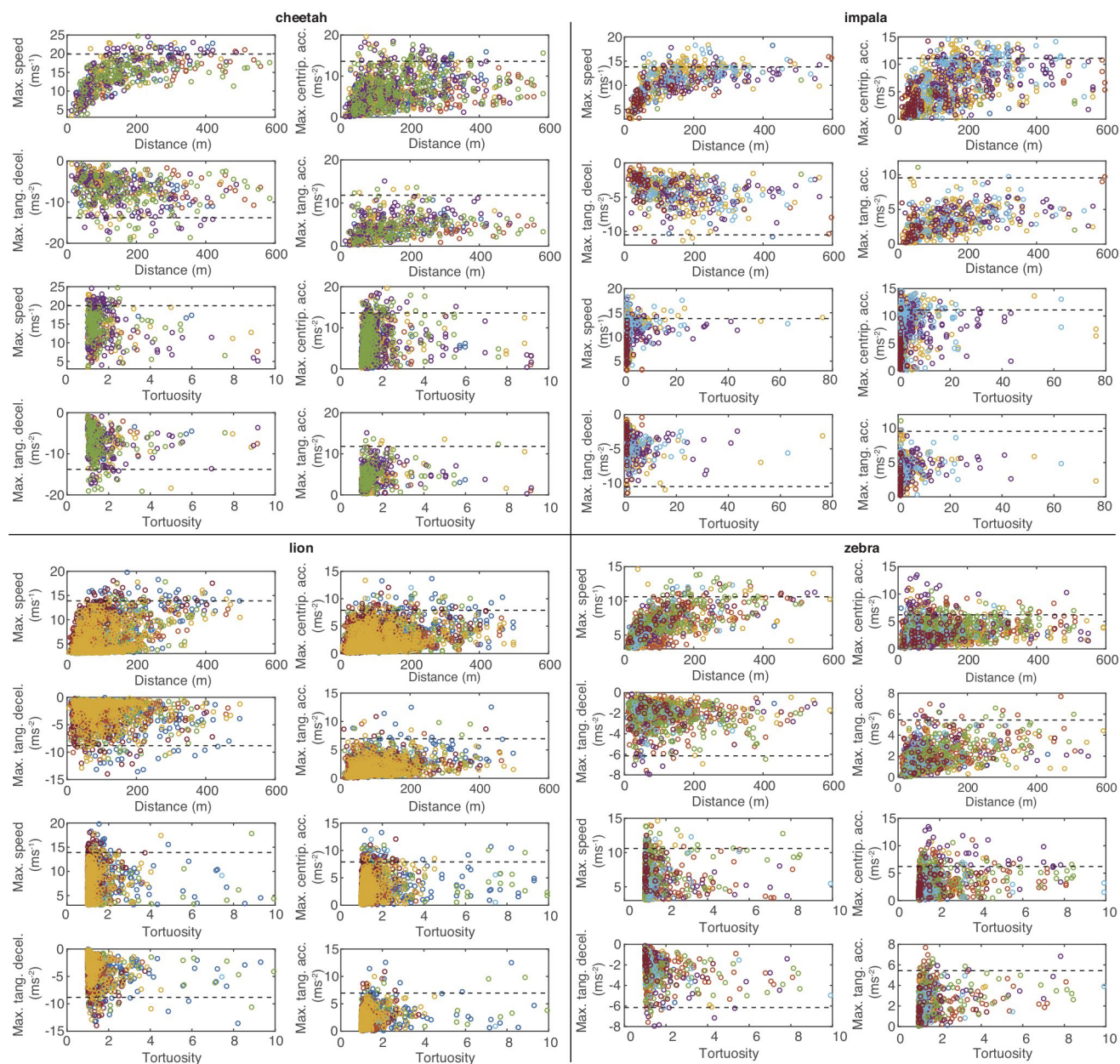
obtained from three temperature jump activations of a single fibre. The curves were fitted (see Methods) to give relative power (power/ $F_{\text{isom}}L_0$  (in  $\text{s}^{-1}$ )) as a function of relative force ( $F/F_{\text{isom}}$ ). Expressing both variables in relative units is important for the curve-fitting process, mainly because the measurements of  $F_{\text{isom}}$  often vary between and within activations; in the example shown in **a** there was a small reduction in  $F_{\text{isom}}$  through the activation at 25°C. **e**, Peak isometric force relative to fibre cross-sectional area (CSA) for fibres from the four species. **f**, Power output relative to fibre volume. There was a distinct subpopulation of low-performance fibres (mostly from lion and zebra) that displayed lower power at a given fibre volume. Fibres with a shortening velocity at peak power of  $<1.35 L_0 \text{ s}^{-1}$  (see also Fig. 2c) were classified as low performance. **g**, Peak power relative to stress at peak power. The low performing fibres also had stress at peak power values that were relatively low—the data points below the thin dashed black line have velocities of shortening  $<1.35 L_0 \text{ s}^{-1}$ . **h**, The variability in stress at peak power was similar across the species tested. **i**, Details about the muscle fibres. Mean ( $\pm$ s.e.m.) mechanical features for single skinned skeletal muscle fibres from biceps femoris of cheetah, lion, impala and zebra. Mean values are also categorized for the predator and prey groups, and further as the high-performing sub-groups of fibres (high-performing fibres had optimal shortening speeds  $>1.35 L_0 \text{ s}^{-1}$ , see main text and Extended Data Fig. 2f, g, i).



### Extended Data Figure 3 | Predator-prey run comparisons.

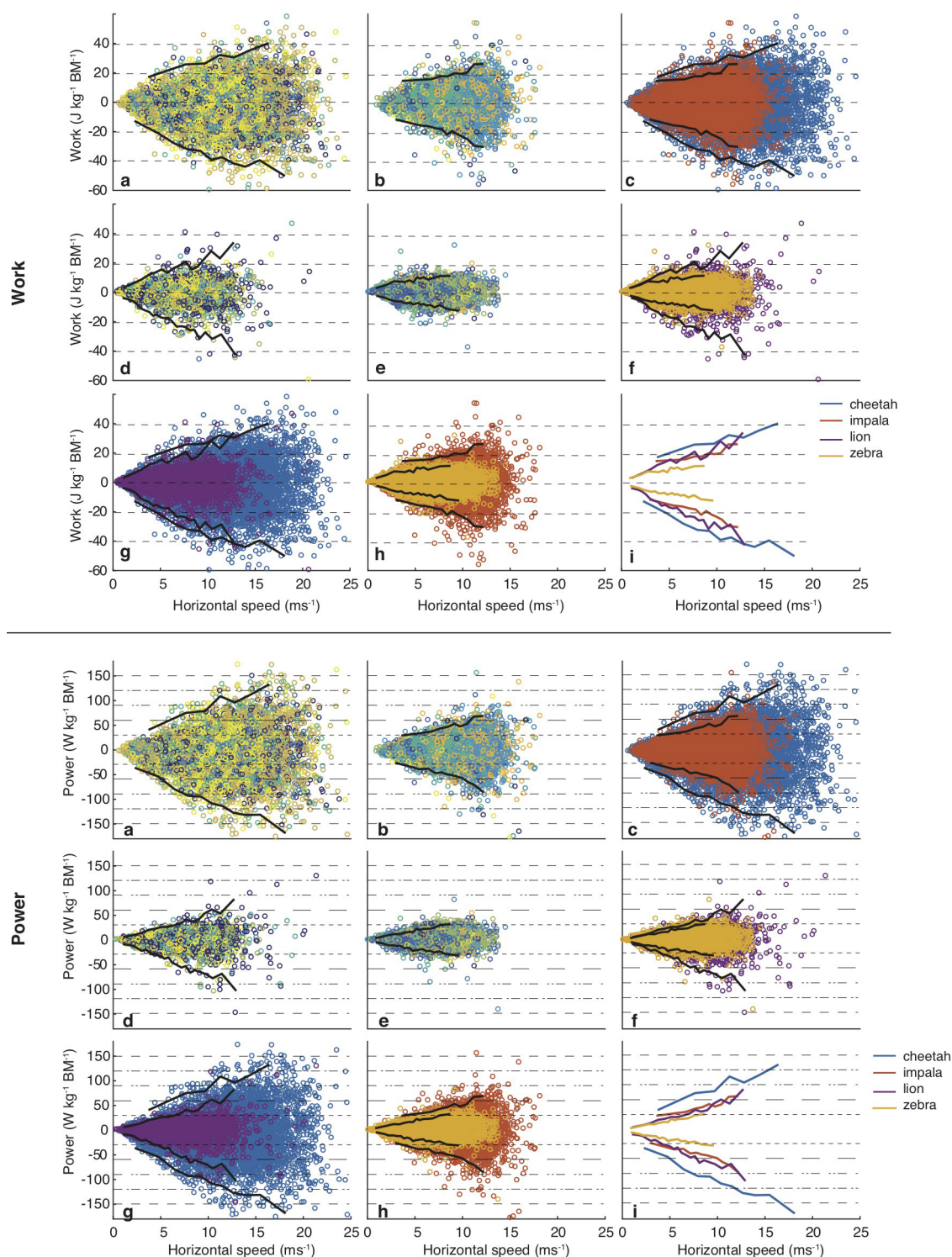
**a, b**, Histograms of all strides (**a**) and extracted non-steady state strides (**b**), for which the cut-off was species-specific based on their kurtosis. The  $x$  axis is the normalized and squared horizontal acceleration in arbitrary units (that is, combined tangential and centripetal). In **b**, on the  $x$  axis, the cut-off is zero and the 98th percentile is one. Cheetah, blue; impala, red; lion, purple; zebra, yellow. Removing steady-state strides delivers a similar distribution tail for all four species. This is critical for deriving an appropriate 98th percentile, as this should be equally representative in all four species. **c**, Histogram of maximum stride parameters recorded in each run (speed, centripetal and tangential acceleration) for each species.

Colour-coded by individuals,  $n$  is the number of runs used for data extraction. One concern of the comparison between predator and prey species is the potential lack of high-performance runs in prey species due to the low number of actual one-on-one chases. However, the distribution of the performance data shows that the cheetah and impala data include a considerably higher proportion of high-performance runs, whereas the lion and zebra dataset includes a large percentage of slower runs. Recognizing that the species differ in run characteristics (motivation, proportion of steady-state versus non-steady-state strides), we removed steady-state strides, based on the species-specific kurtosis, resulting in a comparable distribution in all four species (see Methods).



**Extended Data Figure 4 | Performance metrics separated by individual and species plotted against run distance and against run tortuosity.** Maximum accelerations and speeds were extracted from each run and displayed versus distance covered during the run and versus tortuosity of the run. Tortuosity is the ratio of distance covered in a run to net

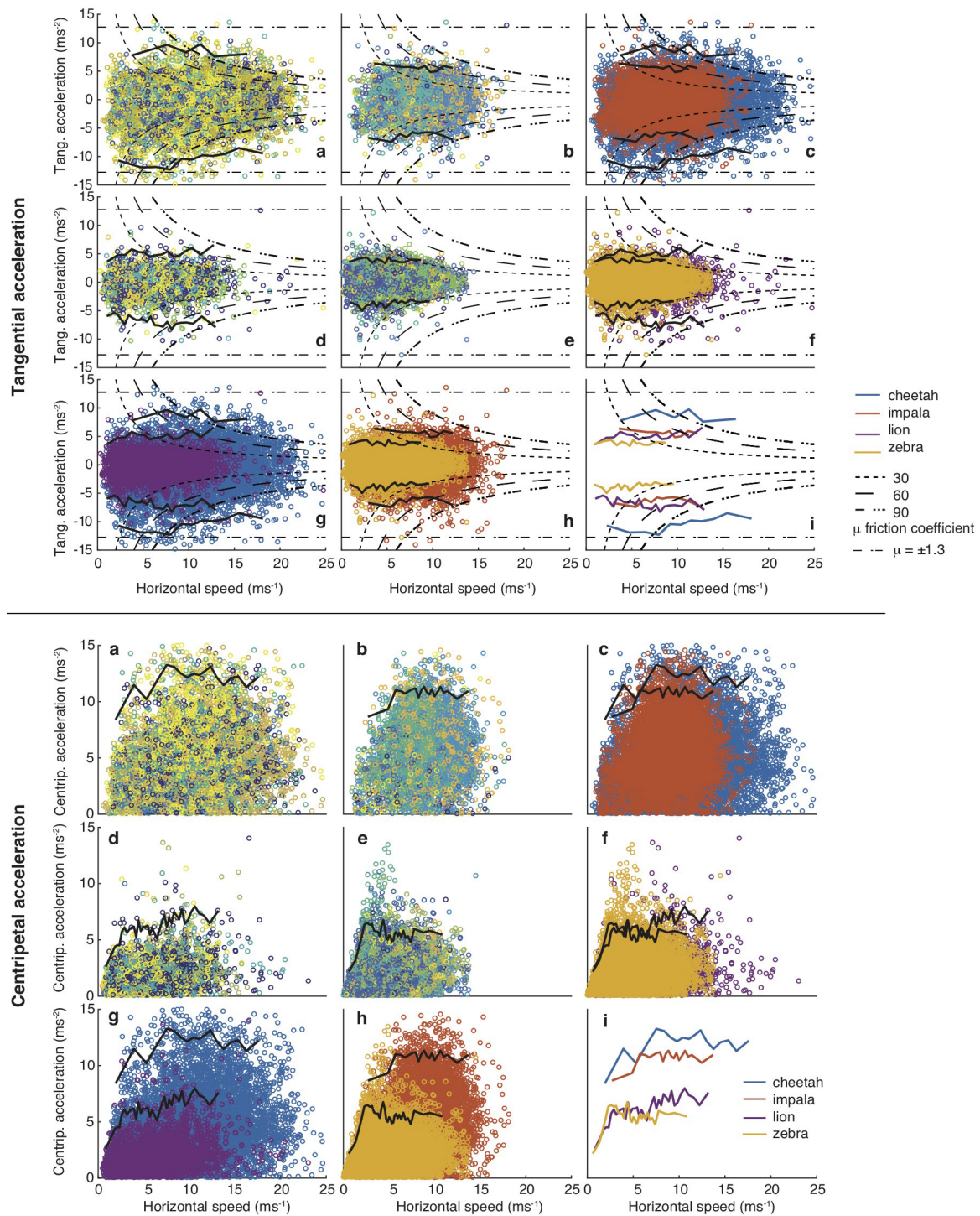
displacement (distance between start and end of the run). Markers are colour-coded per individual, dashed black line maximum values are based on 98th percentile. Number of runs (data points) are given in Extended Data Fig. 1f. Cheetah, 520 runs; impala, 515 runs; lion, 2,726 runs; zebra, 1,801 runs.



### Extended Data Figure 5 | Work and power analysed for each species.

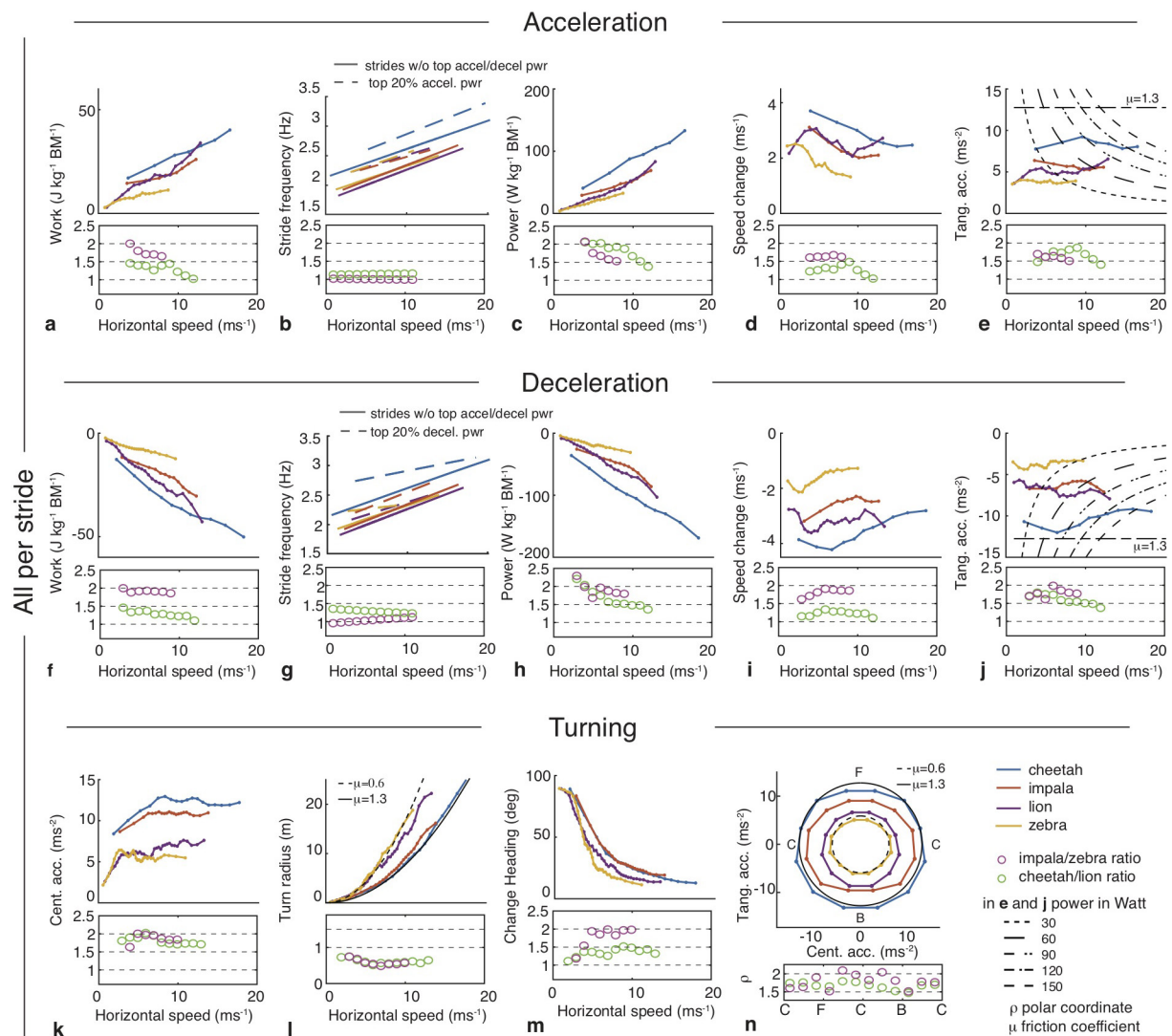
**a**, Cheetah. **b**, Impala. **d**, Lion. **e**, Zebra. **a**, **b**, **d**, **e**, Dots indicate the data points and the line marks the 98th percentile for data in speed bins as shown in Fig. 3a, b, d, e. Markers are colour-coded by individual, the solid black line is the 98th percentile. **c**, **f**, Comparison of the predator-prey pairs. **c**, Cheetah–impala. **f**, Lion–zebra. **g**, Comparison of the predator species (lion–cheetah). **h**, Comparison of the herbivore species (impala–zebra). **i**, The 98th percentile for all four species. **c**, **f**–**i**, Data are colour-coded by species (key is shown in **i**). In all four species maximum negative power was similar to maximum positive power. Muscle stress

can be considerably higher when performing negative work than positive work<sup>70,71</sup> and a 60% higher fascicle power in lengthening (rather than shortening) has been reported<sup>71</sup>, so mass-specific muscle power can be much higher in deceleration<sup>72</sup>. Body geometry relative to the ground reaction force vector or grip may limit the attainable horizontal ground reaction force<sup>15</sup> and the muscles need to be arranged to lengthen while experiencing the large horizontal forces. Many of the propulsive muscles are hip retractors (Extended Data Table 1), which are not configured to resist forward motion.



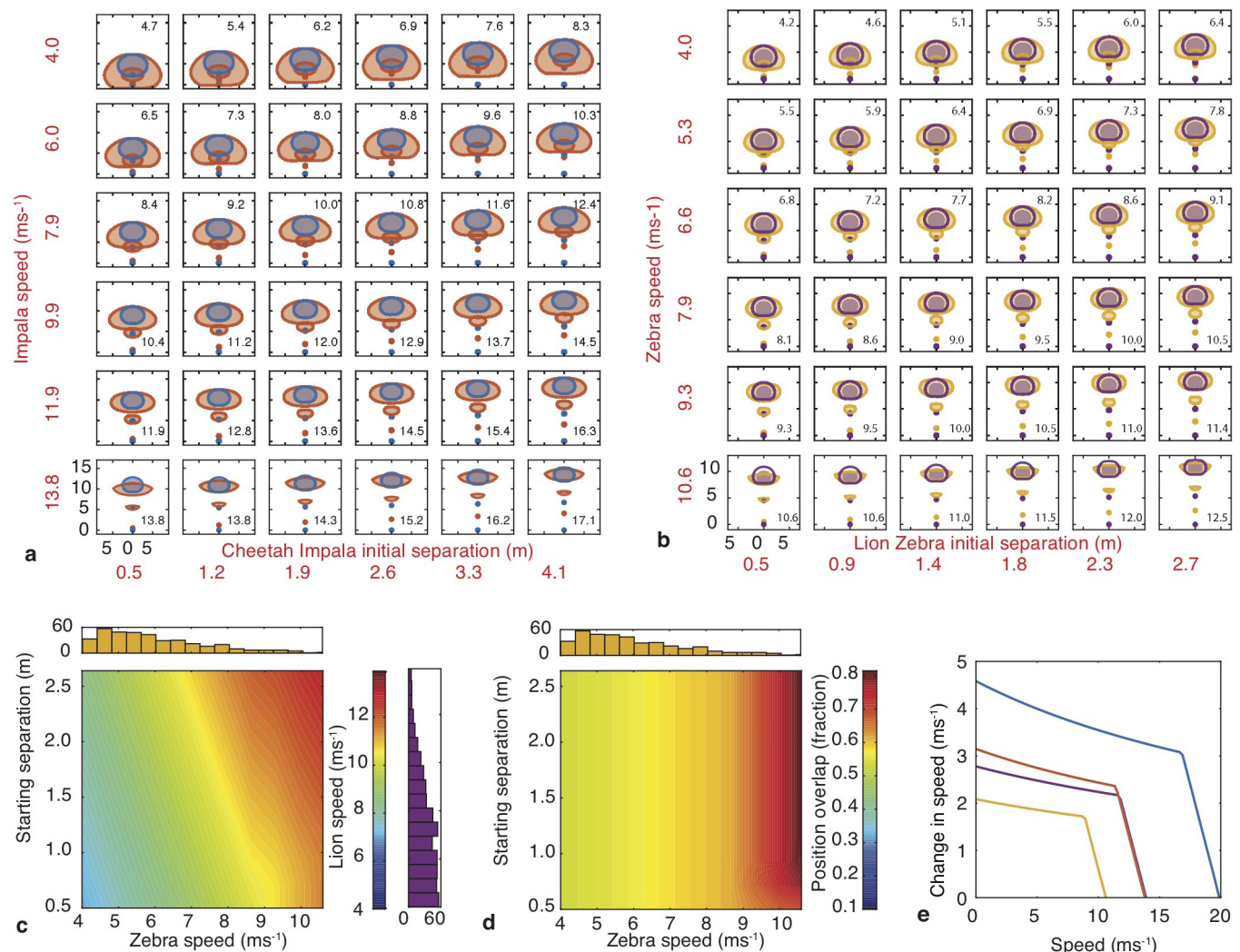
**Extended Data Figure 6 | Tangential and centripetal acceleration analysed for each species. a, Cheetah. b, Impala. d, Lion. e, Zebra.** Dots indicate the data points and the line marks the 98th percentile for data in speed bins as shown in Fig. 3a, b, d, e. Markers are colour-coded by individual, the solid black line is the 98th percentile. **c, f,** Comparison of

the predator-prey pairs. **c,** Cheetah-impala. **f,** Lion-zebra. **g,** Comparison of the predator species (lion-cheetah). **h,** Comparison of the herbivore species (impala-zebra). **i,** The 98th percentile for all four species. **c, f-i,** Data are colour-coded by species (key is shown in **i**).



**Extended Data Figure 7 | Locomotor performance based on stride parameters.** This is the same as Fig. 3, but the ratios compare the two predators and the two prey species. All values are averaged per stride or represent the change over a stride. **a–e**, Acceleration. **a**, Positive net work performed in each stride. **b**, Stride frequency. **c**, Stride power. **d**, Increase in speed per stride. **e**, Forward acceleration with the curved lines representing a mean power of 30, 60, 90, 120 and 150 W. **f–j**, Deceleration. **f–j**, As **a–e** but for decelerating strides. **k–n**, Turning. **k**, Centripetal acceleration. **l**, The relationship between speed and turn

radius with limit lines for a coefficient of friction ( $\mu$ ) of 0.6 and 1.3. **m**, Change in heading versus speed. **n**, Lateral versus tangential acceleration with limits as for  $\mu$ . In **n**, *F* represents pure forward acceleration; *B*, deceleration; and *C*, acceleration to the side. In each panel, there is one line per species that represents the 98th percentile for data in speed bins (bins always include 400 data points, therefore bin width varies). Cheetah, blue; impala, red; lion, purple; zebra, yellow. Bottom, the ratio of that parameter for cheetah to lion (green circle) and impala to zebra (magenta circle) is given for each speed bin.



**Extended Data Figure 8 | Output of model of predator-prey interaction and impact of performance differential on hunt outcome.** See also Fig. 4. **a, b**, Plot showing output of simulation. **a** and **b** are equivalent to Fig. 4d with more subplots for cheetah-impala and lion-zebra, respectively. At the start of simulation, both have initial velocity towards the top of the page and initial separation. After one stride the prey can move to anywhere in the red or yellow ellipse by acceleration in the appropriate direction. Predator velocity remains unchanged, as there is no prey acceleration in the previous stride to react to. Initial positions are shown. Larger red or yellow ellipse perimeter is the area prey can reach after two strides of the chosen maximum acceleration. The blue or purple filled ellipse represents the locations the predator can occupy after its second stride (responding to the prey acceleration observed in first stride). The area of the prey ellipse that is covered by the predator ellipse line is defined as probability of capture. Predator is given a starting speed for each combination of prey speed and initial spacing that maximizes the capture probability. Rows are

different initial prey speeds, values in red to the left of each row. Columns are different initial predator-prey separations at the start of the simulation with values given in red below each column. Scale for all instances is given in the bottom left plot in metres (in black). The inset black numbers in each sub-panel are the initial (optimized for maximum success) predator speeds in  $\text{m s}^{-1}$ . **c**, The optimum lion speed to maximize overlap (hotter colours indicate faster speed, key on the right) as a function of zebra speed ( $x$  axis) and starting separation ( $y$  axis). The histogram above the main plot shows the distribution of actual zebra speed at first turn of 10 degrees or more for each run (same  $x$  axis as the main plot) and the vertical histogram shows distribution of actual lion speed at first turn (scale as for heat bar). **d**, The proportional overlap (capture probability), as a function of zebra initial speed and starting separation. **e**, Modelled capacity for forward acceleration (speed increase per stride) as a function of speed (Extended Data Table 2b). Cheetah, blue; impala, red; lion, purple; zebra, yellow.

**Extended Data Table 1 | Proportion of animal that is locomotor muscle**

| Species                     | n                 | Provenance | Animal mass (kg) | Hind leg muscle | Front leg muscle | Spine/back muscle | Total locomotor muscle |
|-----------------------------|-------------------|------------|------------------|-----------------|------------------|-------------------|------------------------|
| Greyhound <sup>51,52</sup>  | 7 front<br>6 hind | R<br>R     | 31.8             | 18.80%          | 18.40%           | ---               | ---                    |
| Greyhound <sup>53</sup>     | 9                 | R          | 27.3             | ---             | ---              | 9.70%             | ---                    |
| Cheetah <sup>54,55</sup>    | 8                 | Z          | 33.1             | 18.30%          | 14.20%           | ---               | ---                    |
| Impala                      | 5                 | W          | 49.7             | 17.50%          | 11.30%           | ---               | ---                    |
| Horse <sup>56</sup>         | 7                 | D          | 510              | 19.40%          | ---              | ---               | ---                    |
| Horse <sup>57</sup>         | 6                 | D          | 503              | ---             | 6.60%            | ---               | ---                    |
| Arabian horse <sup>58</sup> | 6                 | D          | 383              | 10.20%*         | ---              | ---               | ---                    |
| Quarterhorse <sup>58</sup>  | 6                 | D          | 457              | 12.70%*         | ---              | ---               | ---                    |
| Hare <sup>59,60</sup>       | 8                 | W          | 3.45             | 16.30%          | 9.30%            | 8.90%             | 34.50%                 |
| Ostrich <sup>61</sup>       | 11                | F          | 105              | 33.70%          | ---              | ---               | ---                    |
| Ostrich <sup>62</sup>       | 2                 | F          | 100              | 30.29%          | ---              | ---               | ---                    |
| Lion <sup>63,64</sup>       | 1                 | Z          | 133              | 12.48%          | 14.30%           | ---               | ---                    |
| Beef Cattle <sup>65</sup>   | ---               | F          | 544.3            | 28.00%          | 8.60%            | ---               | ---                    |
| Domestic Cat <sup>66</sup>  | ---               | D          | 3.25             | 14.00%          | ---              | ---               | ---                    |
| Mix-Breed Dog <sup>67</sup> | 1                 | D          | 23               | 10.40%          | ---              | ---               | ---                    |

Data are taken from published values for athletic species and number of animals dissected given when reported<sup>51–67</sup>. Provenance: wild (W), racing/competition (R), zoo (Z), farmed (F), domestic (D). Complete datasets, including all muscles and the animal mass, are sparse and many of the ones summarized are from our own group. In some of the studies, the animals would be sedentary or have died from other causes, so it is likely that wild animals have more muscle<sup>9</sup>. It was previously reported<sup>68</sup> that skeletal muscle as a fraction of body mass is 53% for thoroughbred racehorses and 44% for other horses. All values are for muscle from both limbs as a percentage of total body mass.

\*Only 11 muscles reported.

Extended Data Table 2 | Performance parameters

**a**

| Performance parameter                             | Predator | Herbivore |
|---|----------|-----------|
| Pos. tangential acceleration ( $\text{ms}^{-2}$ ) | -0.5     | -0.21     |
| Neg. tangential acceleration ( $\text{ms}^{-2}$ ) | -0.42    | -0.28     |
| Pos. work ( $\text{Jkg}^{-1}\text{BM}^{-1}$ )     | -0.17    | -0.45     |
| Neg. work ( $\text{Jkg}^{-1}\text{BM}^{-1}$ )     | 0.16     | 0.48      |
| Pos. power ( $\text{Wkg}^{-1}\text{BM}^{-1}$ )    | -0.47    | -0.41     |
| Neg. power ( $\text{Wkg}^{-1}\text{BM}^{-1}$ )    | 0.5      | 0.55      |

**b**

| Species | Speed ( $\text{ms}^{-1}$ ) | Centrip. accel ( $\text{ms}^{-2}$ ) | Pos. tang. accel ( $\text{ms}^{-2}$ ) | Neg. tang. accel ( $\text{ms}^{-2}$ ) | Pos. work ( $\text{Jkg}^{-1}\text{BM}$ ) | Neg. work ( $\text{Jkg}^{-1}\text{BM}$ ) | Pos. power ( $\text{Wkg}^{-1}\text{BM}$ ) | Neg. power ( $\text{Wkg}^{-1}\text{BM}$ ) |
|---------|----------------------------|-------------------------------------|---------------------------------------|---------------------------------------|--|--|---|---|
| Cheetah | 19.9                       | 12.3                                | 8.3                                   | -10.6                                 | 33.3                                     | -39                                      | 109                                       | -114                                      |
| Impala  | 13.8                       | 10.9                                | 5.7                                   | -6.3                                  | 22.1                                     | -24                                      | 59  | -60                                       |
| Lion    | 13.9                       | 6.5                                 | 5.2                                   | -7                                    | 20.5                                     | -24                                      | 48  | -56                                       |
| Zebra   | 10.6                       | 5.6                                 | 3.9                                   | -3.8                                  | 9.9                                      | -9                                       | 24  | -22                                       |

**a**, log–log slope of performance parameter versus mass for the two herbivore species and the two predator species. Stride values were extracted to represent species' performance and evaluated versus body mass to explore whether the performance difference between small and large was concomitant with effects reported across a broad animal size range<sup>69</sup>. Parameters with increasing values (positive and negative work and power) were represented by maximum values whereas for parameters that plateaued (positive and negative tangential acceleration, centripetal acceleration), an average value was calculated. The slope of the logarithmic coordinates (log–log slope of performance parameter versus mass) was calculated for the two predators and two herbivores. The relationship is generally consistent in predators and in prey, with most parameters dropping with increasing size, but this does not provide an explanation for the magnitude of the differences seen, as most parameters would scale weakly with animal size. **b**, Maximum (98%) values for stride parameters for all species. Maximum values were determined using the 98th percentile (after species-specific steady-state strides were removed from the data, positive and negative data were calculated separately). These are the parameters used in the model.

# HER kinase inhibition in patients with HER2- and HER3-mutant cancers

David M. Hyman<sup>1</sup>, Sarina A. Piha-Paul<sup>2</sup>, Helen Won<sup>1</sup>, Jordi Rodon<sup>3</sup>, Cristina Saura<sup>3</sup>, Geoffrey I. Shapiro<sup>4</sup>, Dejan Juric<sup>5</sup>, David I. Quinn<sup>6</sup>, Victor Moreno<sup>7</sup>, Bernard Døger<sup>7</sup>, Ingrid A. Mayer<sup>8</sup>, Valentina Boni<sup>9</sup>, Emiliano Calvo<sup>9</sup>, Sherene Loi<sup>10</sup>, Albert C. Lockhart<sup>11</sup>, Joseph P. Erinjeri<sup>1</sup>, Maurizio Scaltriti<sup>1</sup>, Gary A. Ulaner<sup>1</sup>, Juber Patel<sup>1</sup>, Jiabin Tang<sup>1</sup>, Hannah Beer<sup>1</sup>, S. Duygu Selcuklu<sup>1</sup>, Aphrothiti J. Hanrahan<sup>1</sup>, Nancy Bouvier<sup>1</sup>, Myra Melcer<sup>1</sup>, Rajmohan Murali<sup>1</sup>, Alison M. Schram<sup>1</sup>, Lillian M. Smyth<sup>1</sup>, Komal Jhaveri<sup>1</sup>, Bob T. Li<sup>1</sup>, Alexander Drilon<sup>1</sup>, James J. Harding<sup>1</sup>, Gopa Iyer<sup>1</sup>, Barry S. Taylor<sup>1</sup>, Michael F. Berger<sup>1</sup>, Richard E. Cutler Jr<sup>12</sup>, Feng Xu<sup>12</sup>, Anna Butturini<sup>12</sup>, Lisa D. Eli<sup>12</sup>, Grace Mann<sup>12</sup>, Cynthia Farrell<sup>12</sup>, Alshad S. Lalani<sup>12</sup>, Richard P. Bryce<sup>12</sup>, Carlos L. Arteaga<sup>8</sup>, Funda Meric-Bernstam<sup>2</sup>, José Baselga<sup>1</sup> & David B. Solit<sup>1</sup>

**Somatic mutations of *ERBB2* and *ERBB3* (which encode HER2 and HER3, respectively) are found in a wide range of cancers. Preclinical modelling suggests that a subset of these mutations lead to constitutive HER2 activation, but most remain biologically uncharacterized. Here we define the biological and therapeutic importance of known oncogenic HER2 and HER3 mutations and variants of unknown biological importance by conducting a multi-histology, genomically selected, ‘basket’ trial using the pan-HER kinase inhibitor neratinib (SUMMIT; clinicaltrials.gov identifier NCT01953926). Efficacy in HER2-mutant cancers varied as a function of both tumour type and mutant allele to a degree not predicted by preclinical models, with the greatest activity seen in breast, cervical and biliary cancers and with tumours that contain kinase domain missense mutations. This study demonstrates how a molecularly driven clinical trial can be used to refine our biological understanding of both characterized and new genomic alterations with potential broad applicability for advancing the paradigm of genome-driven oncology.**

Genomic profiling of human cancers has identified recurrent somatic mutations of HER2 (encoded by *ERBB2*) and HER3 (*ERBB3*), typically occurring in the absence of gene amplification<sup>1–3</sup>. Mutations in HER2 are clustered in the extracellular, transmembrane and kinase domains. Unlike other mutant oncogenes, such as *BRAF* or *KRAS*, no single mutant allele predominates and the precise distribution of mutations varies by tumour type<sup>4</sup>. By contrast, HER3 mutations cluster primarily in the extracellular domain and to a lesser extent in the kinase domain. Although HER2 and HER3 mutations are found in a wide variety of cancers, their overall prevalence does not exceed 10% in any individual tumour type, and the rate is more typically less than 5% for HER2 and less than 1% for HER3.

Biological modelling has yielded conflicting findings as to the functional consequences of HER2 and HER3 mutations. Substantial data suggest that a subset of these mutations induce ligand-independent constitutive HER2 receptor signalling and promote oncogenesis<sup>5–7</sup>. The mechanism of these oncogenic effects seems to differ by variant, with some causing enhanced HER2 kinase activity and others causing receptor dimerization<sup>5,8</sup>. Mutations in HER3, which in its wild-type configuration has impaired kinase function, seem to rely on wild-type HER2 to exert its oncogenic effects<sup>7</sup>. Most preclinical data that explore the functional consequences of HER2 and HER3 mutations have been generated using engineered models that overexpress the mutation, and thus the results may be confounded by the known oncogenic effects of HER2 overexpression. Further enforcing the potential importance of this confounding variable, models of HER2 mutation generated by gene-editing techniques have failed to demonstrate a malignant phenotype in the absence of mutations in other oncogenes such as *PIK3CA*<sup>9</sup>.

Given the considerable diversity of HER2 and HER3 mutations, as well as the challenge of generating preclinical models that recreate their true biology in human cancers, we sought to define the therapeutic importance of HER2 and HER3 mutations by conducting SUMMIT—a global, multicentre, multi-histology basket trial in patients with tumours that contain these mutations (Extended Data Fig. 1). Patients were treated with neratinib, an irreversible pan-HER tyrosine kinase inhibitor, which potently inhibits the growth of HER2-mutant tumours in preclinical models<sup>5</sup>. Tumour tissue and plasma were collected to facilitate the detailed genomic characterization of patients. Here we present the results of this study, with a focus on the insights it provides into the biological and therapeutic importance of HER2 and HER3 mutations in patients with cancer.

## Patient and mutation characteristics

Baseline patient demographics are shown in Table 1 and Extended Data Table 1. In total, 141 patients (125 with HER2-mutant tumours, 16 with HER3-mutant tumours) received neratinib treatment. These patients were diagnosed with 1 out of 21 unique cancer types, the most common being breast, lung, bladder and colorectal cancer (61% of patients treated). As has been seen in other basket studies<sup>10,11</sup>, we identified and enrolled several orphan tumour types including cancers of the biliary tract, salivary gland, small bowel and vagina, as well as extramammary Paget’s disease (in aggregate, 13% of all patients). Patients tended to be heavily pretreated with approximately half having received at least three previous lines of systemic therapy.

Enrolled patients had 31 unique HER2 and 11 unique HER3 mutations (Extended Data Fig. 2). The most frequent HER2 mutations

<sup>1</sup>Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>2</sup>University of Texas, MD Anderson Cancer Center, Houston, Texas, USA. <sup>3</sup>Vall d’Hebron University Hospital, Vall d’Hebron Institute of Oncology (VHIO), Barcelona, Spain. <sup>4</sup>Dana-Faber Cancer Institute, Boston, Massachusetts, USA. <sup>5</sup>Massachusetts Hospital Cancer Center, Boston, Massachusetts, USA. <sup>6</sup>USC Norris Comprehensive Cancer Center, Los Angeles, California, USA. <sup>7</sup>START Madrid Fundación Jiménez Díaz, Madrid, Spain. <sup>8</sup>Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA. <sup>9</sup>START Madrid, Centro Integral Oncológico Clara Campal (CIOCC), Madrid, Spain. <sup>10</sup>Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. <sup>11</sup>Washington University in St. Louis School of Medicine, St Louis, Missouri, USA. <sup>12</sup>Puma Biotechnology Inc., Los Angeles, California, USA.

**Table 1 | Patient demographics**

| Patient characteristic   | HER2 mutant (n = 125) | HER3 mutant (n = 16) | Total (n = 141) |
|--|-----------------------|----------------------|-----------------|
| <b>Age, years</b>  |                       |                      |                 |
| Median (range)   | 61 (30–83)            | 66 (39–82)           | 61 (30–83)      |
| <65 years, n (%)   | 81 (64.8)             | 7 (43.8)             | 88 (62.4)       |
| ≥65 years, n (%)   | 44 (35.2)             | 9 (56.3)             | 53 (37.6)       |
| <b>Sex, n (%)</b>  |                       |                      |                 |
| Female   | 80 (64.0)             | 12 (75.0)            | 92 (65.2)       |
| Male   | 45 (36.0)             | 4 (25.0)             | 49 (34.8)       |
| <b>ECOG performance status, n (%)</b>                          |                       |                      |                 |
| 0  | 37 (29.6)             | 1 (6.3)              | 38 (27.0)       |
| 1  | 83 (66.4)             | 12 (75.0)            | 95 (67.4)       |
| 2  | 5 (4.0)               | 3 (18.8)             | 8 (5.7)         |
| <b>Previous systemic treatment lines, n (%)</b>                |                       |                      |                 |
| Any  | 121 (96.8)            | 16 (100)             | 137 (97.2)      |
| 1  | 33 (26.4)             | 1 (6.3)              | 34 (24.1)       |
| 2  | 30 (24.0)             | 11 (68.8)            | 41 (29.1)       |
| ≥3   | 58 (46.4)             | 4 (25.0)             | 62 (44.0)       |
| <b>Median time from metastasis to enrolment, years (range)</b> | 1.02 (0.0–15.0)       | 1.13 (0.3–4.5)       | 1.03 (0.0–15.0) |
| <b>Tumour type, n (%)</b>                                      |                       |                      |                 |
| Lung   | 26 (20.8)             | 0 (0)                | 26 (18.4)       |
| Breast   | 25 (20.0)             | 0 (0)                | 25 (17.7)       |
| Bladder  | 16 (12.8)             | 2 (12.5)             | 18 (12.8)       |
| Colorectal   | 12 (9.6)              | 5 (31.3)             | 17 (12.1)       |
| Biliary tract  | 9 (7.2)               | 2 (12.5)             | 11 (7.8)        |
| Endometrial  | 7 (5.6)               | 1 (6.3)              | 8 (5.7)         |
| Cervical   | 5 (4.0)               | 0 (0)                | 5 (3.5)         |
| Gastroesophageal   | 5 (4.0)               | 2 (12.5)             | 7 (5.0)         |
| Ovarian  | 4 (3.2)               | 1 (6.3)              | 5 (3.5)         |
| Other  | 16 (12.8)             | 3 (18.8)             | 19 (13.5)       |

were S310, L755, Y772\_A775dup and V777 alleles. The HER2 kinase domain was most commonly mutated (66%), followed by the extracellular (26%) and transmembrane/juxtamembrane (8%) domains. The anticipated relationships between the mutated HER2 domain and tumour type were observed, with extracellular domain mutations predominant in bladder cancer, kinase domain missense mutations in breast and colon cancer, and kinase domain insertions in lung cancer<sup>4</sup>. Missense mutations were the most common class of genomic alteration (74%), followed by in-frame insertions (22%), the latter exclusively affecting the kinase domain. Two tumours contained HER2 insertions/deletions and one an in-frame kinase domain-retaining fusion (*GRB7-ERBB2*)<sup>12,13</sup>. HER3 mutations were all missense variants and clustered in the extracellular furin-like and receptor domains. In total, 87% (109 out of 125) of HER2 and 75% (12 out of 16) of HER3 mutations were at positions now known to be mutational hotspots<sup>4</sup>. This pattern of HER2 and HER3 mutations was comparable to the spectrum of non-truncating HER2 and HER3 mutations observed in previously published genomic landscape studies, including The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC)<sup>4</sup>, although HER2 V777L and Y772\_A755dup were more common in our study cohort (13.6% versus 5.3% and 12.0% versus 2.7%, respectively; Extended Data Fig. 3).

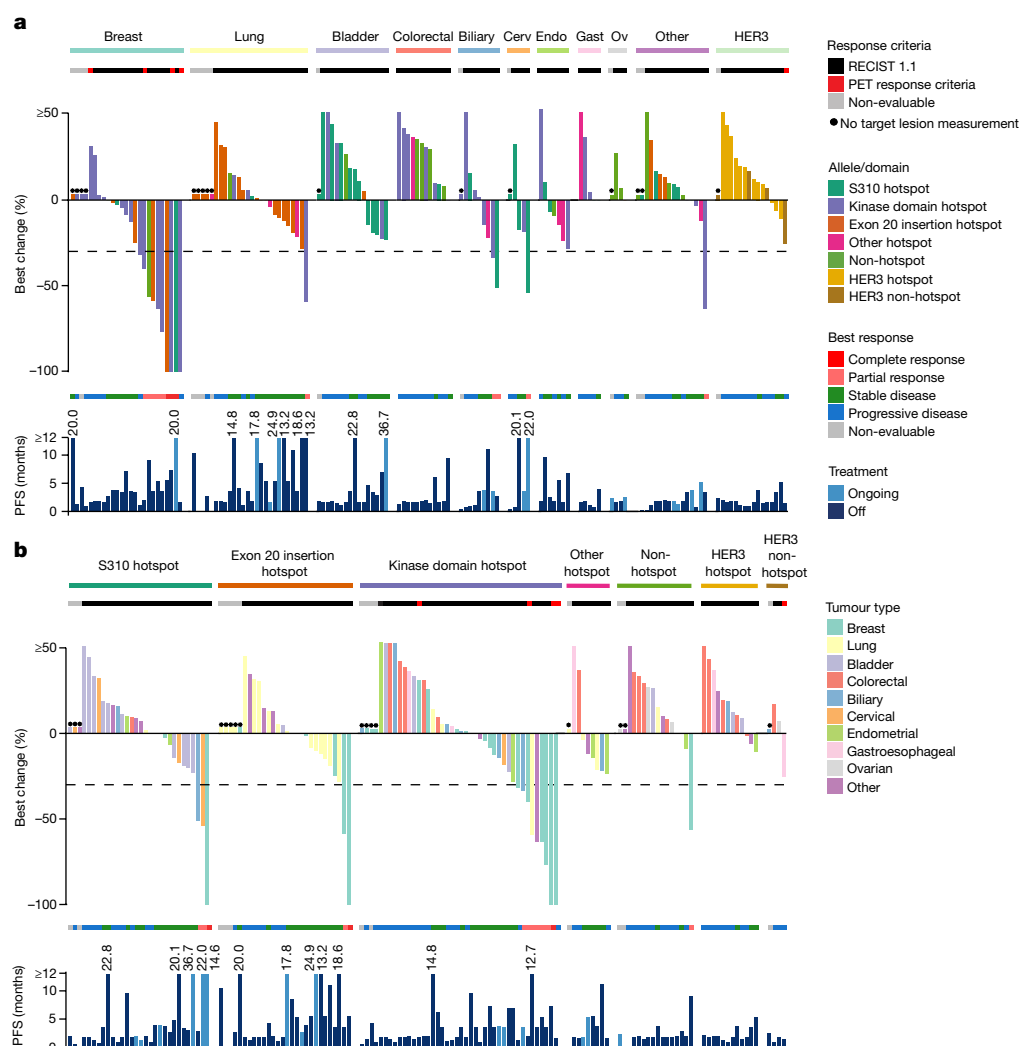
## Treatment outcomes

When stratified by tumour type, we observed responses to neratinib in patients with HER2-mutant breast, non-small-cell lung, cervical, biliary and salivary cancers, which led to expanded enrolment in several of these tumour types (Fig. 1a, Extended Data Table 1). Neratinib exhibited the greatest degree of activity in patients with breast cancer ( $n = 25$  total, objective response rate at week 8 (ORR<sub>8</sub>) 32%, 95% confidence interval 15–54%), with responses observed in patients with missense mutations involving the extracellular and kinase domains, as well as insertions in the kinase domain. All patients with breast cancer

were classified as HER2-negative (non-amplified) at the time of enrolment as per established guidelines<sup>14</sup>. Responses were observed in both oestrogen receptor-positive (30%, 6 out of 20) and -negative (40%, 2 out of 5) tumours. Overall, these breast cancer data are generally consistent with a previous report<sup>15</sup>. In patients with lung cancer ( $n = 26$ ), in which insertions in exon 20 predominate, we observed only one objective response. Of note, HER2 exon 20 insertions are paralogous of EGFR exon 20 insertions, which are resistant to first- and second-generation EGFR tyrosine kinase inhibitors<sup>16</sup>. Notably, the only patient with lung cancer to achieve a response evaluation criteria in solid tumours (RECIST) response had a kinase domain missense mutation (L755S). Despite the low response rate, the median progression-free survival in recurrent lung cancer was 5.5 months, with 6 patients remaining on therapy for more than 1 year, which compares favourably to second-line chemotherapy and immune checkpoint inhibitors<sup>17</sup>, suggesting that neratinib may have a positive effect on the natural history of this disease. Responses were also observed in biliary and cervical cancers, and enrolment is ongoing in these cohorts to define this activity better. No responses were observed in bladder cancer ( $n = 16$ ) or colorectal cancer ( $n = 12$ ), suggesting lineage-dependent resistance to single-agent pan-HER kinase inhibition in these tumour types. In summary, among the HER2-mutant cohorts, breast cancer met the primary endpoint for efficacy, whereas lung, colorectal and bladder cancers did not. For the remaining tumour-specific cohorts, enrolment is continuing and they have therefore not undergone final efficacy analysis. Despite preclinical data to suggest that HER3 mutations can be oncogenic drivers, no responses to neratinib were observed in patients with HER3-mutant tumours.

When stratified by mutant allele, responses were observed in patients with tumours containing HER2 S310, L755, V777, G778\_P780dup and Y772\_A775dup mutations (Fig. 1b). Among patients with HER2 kinase domain hotspot missense mutations ( $n = 42$ ), responses were noted in four unique tumour types (breast, biliary, lung and salivary gland). By allele, we observed responses in several kinase domain mutants including L755S ( $n = 4$ ), V777L ( $n = 4$ ) and L869R ( $n = 1$ ). In patients with HER2 hotspot extracellular domain mutations (S310,  $n = 30$ ), responses were observed in breast, cervical and biliary cancers ( $n = 1$  for each), but not in bladder cancer, the cancer type in which these mutations predominate. Similarly, in patients with HER2 exon 20 insertions ( $n = 28$ ), responses were observed in two patients with breast cancer, but none were seen in patients with lung cancer, in which this class of alteration is most common. In exon 20 insertions, preservation of glycine at the 770 position, which seems to facilitate binding of covalent HER kinase inhibitors such as neratinib, did not predict for response as previously suggested by preclinical modelling<sup>18</sup> (Extended Data Fig. 4). Similarly, the number of amino acids involved in the insertion did not seem to predict outcome, with responses observed in patients with both 3 (G788\_P780dup) and 4 (Y722\_A755dup) amino acid insertions. Finally, among the 15 patients with HER2 mutations not known to be hotspots, only one responded to neratinib. Notably, this response occurred in a patient with breast cancer and a complex insertion/substitution (L755\_E757delinsS), which, to our knowledge, has not been observed previously. Although this case illustrates that the tumours of some patients may be addicted to truly private oncogenic drivers (those arising in only a single patient), it is also noteworthy that this insertion occurs in a domain that is the target of recurrent insertions. The absence of clinical activity in the remaining 14 patients with cancers with non-hotspot mutations suggests that, although the recurrence of a mutation in HER2 is insufficient to define it as sensitizing to a HER2 kinase inhibitor, the absence of recurrence (that is, mutations that do not occur at hotspot positions) provides circumstantial evidence that the alteration is unlikely to be a driver.

Although the overall numbers of patients in each subgroup preclude formal statistical comparison, integrating efficacy, mutational and lineage data, we observed that clinical benefit from neratinib therapy appeared to vary as a function of both mutational and disease context



**Figure 1 | Individual treatment outcome and response for 141 patients grouped by tumour cohort and mutant allele/domain. a, b, Top, percentage best change from baseline in the target lesion assessed by the appropriate response criteria (RECIST version 1.1 or PET). Each bar is colour coded according to its mutation allele/domain, for patients grouped by tumour cohort (a), or tumour type, for patients grouped by mutant allele/domain (b). Middle, best overall response. Bottom, progression-free survival (PFS), colour-coded by treatment status. \*Non-evaluable. Cerv, cervical; endo, endometrial; gastro, gastroesophageal; ov, ovarian; PET, positron-emission tomography.**

(Fig. 2). In tumour types sensitive to neratinib therapy, such as breast, biliary and cervical cancers, responses were collectively observed across all types and classes of HER2 mutations. By contrast, in lung cancer, a tumour type that exhibits modest sensitivity to neratinib, response was limited to a patient with a HER2 kinase domain missense mutation—a class of mutation with greater *in vitro* sensitivity to neratinib<sup>5</sup>. Finally, in tumour types with intrinsic lineage-based resistance to neratinib, such as bladder and colorectal cancers, responses were not observed regardless of the HER2 mutation, type or class.

## Safety

All patients received neratinib with mandatory anti-diarrhoeal prophylaxis. With this regimen, the rate of grade 3 diarrhoea was 22% (Extended Data Table 2), consistent with previous experience<sup>19</sup>. Among patients who developed grade 3 diarrhoea, the median time to onset was 10 days and the median duration of the diarrhoea episode was 2 days. Patients were typically managed with dose interruption and reduction, with only 2.8% permanently discontinuing therapy owing to diarrhoea. The remainder of adverse events were predominantly low-grade.

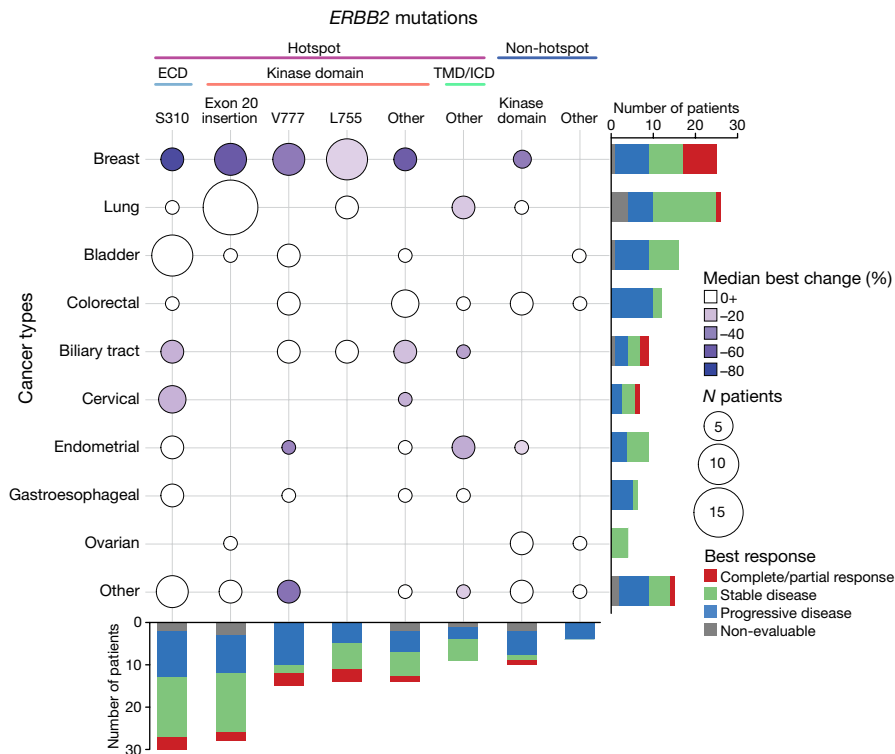
## Central confirmation of HER2 and HER3 mutations

There is active debate within the cancer research community as to whether central confirmation of mutational status before study entry is optimal for determining trial eligibility for precision medicine studies. To define the reproducibility of local mutational testing, DNA from archival formalin-fixed paraffin-embedded tumour and plasma

samples were re-sequenced (see Methods). A total of 33 patients (26 HER2-mutant, 7 HER3-mutant) were excluded from this concordance analysis because the local test used was the same as the central tumour assay being evaluated. Of the remaining 99 patients with HER2 mutations, adequate material for tumour genomic testing was unobtainable for 26 patients. Overall, concordance in the remaining patients based on central tumour and/or plasma sequencing was 95% (69 out of 73), with 38 patients assessed by tissue and plasma, 14 by tissue alone, and 21 by plasma alone. Central testing identified one locally reported mutation (V773M) as a germline polymorphism and this patient, with renal cell carcinoma, had progressive disease at first scan. Central testing in the four cases in which the HER2 mutation could not be confirmed passed all quality-control metrics, but in two patients the testing was performed on material collected at least three years after the tissue used for local testing, raising the possibility that tumour heterogeneity was involved in the discordance. None of the patients with discordant HER2 results responded to neratinib, and their median progression-free survival was only 43 days (range: 5–58 days). Among the 9 patients eligible for concordance testing with HER3 mutations, tumour tissue was available for central sequencing in 8 patients, and overall concordance was 75% (6 out of 8).

## Genomic modifiers of response

Given the variability of treatment response, even among patients with the same tumour lineage and HER2-mutant allele, we sought to identify other genomic modifiers of response through broader genomic characterization of tumour-derived DNA (see Methods). First, we



**Figure 2 | Integrated efficacy by tumour type and HER2 allele/domain.**

The y axis represents the tumour types, and the x axis represents the mutated allele/domain and hotspot status. The hotspot mutations are further broken down into the various domains. The size of the circle is proportional to the count of the tumour type and allele/domain; the

colour of the circle reflects the median percentage best change in the target lesions (any zero or positive median change is indicated in white). The stacked bars represent the best overall response for the tumour type or domain/allele, as indicated in the key. ECD, extracellular domain; ICD, intracellular domain; TMD, transmembrane domain.

explored the relationship between *ERBB2* amplification and outcome, as this is a well-established predictor of response to HER2-targeted therapies in patients lacking HER2 mutations. In total, 17% of patients (15 out of 86) had concurrent HER2 mutations and gene amplification. Amplifications preferentially targeted the mutant allele locus (86%, 12 out of 14 evaluable). Using a dichotomous definition of clinical benefit (stable disease or partial response lasting at least 24 weeks), *ERBB2* amplification did not correlate with outcome ( $P=0.50$ ; Fig. 3), suggesting that in the presence of HER2 mutations, amplification may not confer additional sensitivity to irreversible HER kinase inhibitors. We also explored the relationship of *ERBB2* mutation clonality on outcomes. In the 74 patients with adequate material to allow definitive assessment of *ERBB2* mutant clonality, the mutation was clonal in 95% (70 out of 74; Extended Data Fig. 5a). None of four patients with a subclonal *ERBB2* mutation achieved clinical benefit.

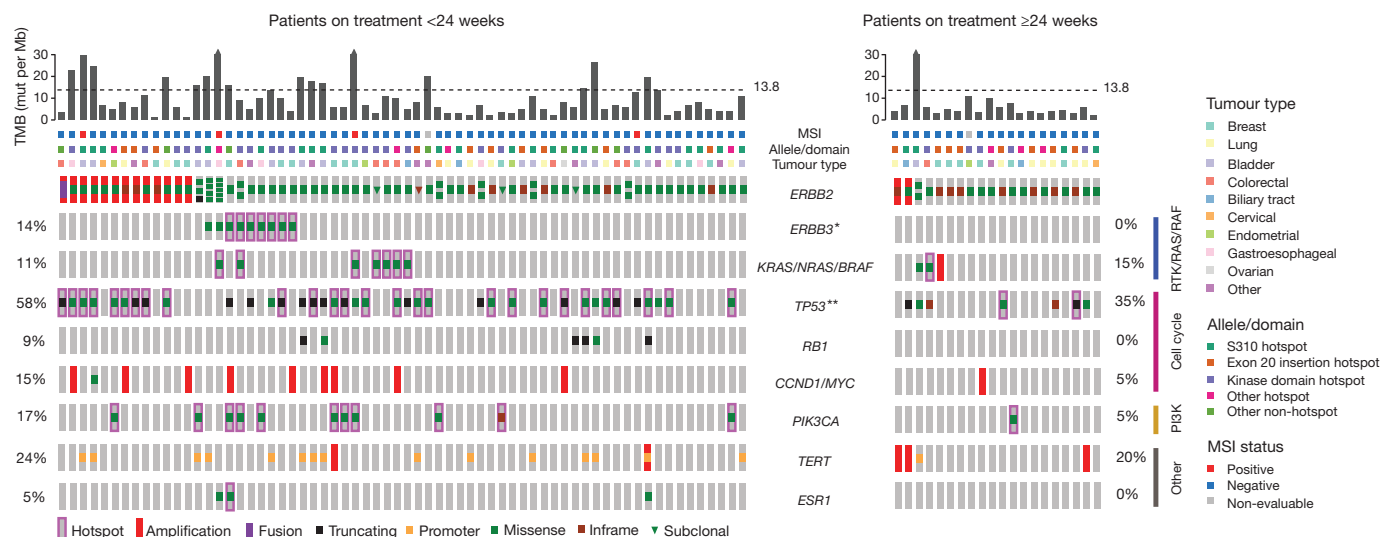
Hypothesizing that tumours with an increased tumour mutational burden (TMB) might be more likely to acquire HER2 mutations without developing oncogenic dependence (that is, passenger mutations), we evaluated whether overall TMB status affected outcome. Using a previously validated cut-off ( $\geq 13.8$  non-synonymous mutations per megabase of DNA<sup>2</sup>), 20% of patients (17 out of 86) met criteria for a high TMB. In total, 24% of patients (16 out of 66) without clinical benefit versus 5% of patients (1 out of 20) with benefit met criteria for a high TMB, a trend that did not reach statistical significance ( $P=0.10$ ).

Next, we evaluated whether the pattern of co-mutations affected clinical benefit in the subset of patients where broader profiling was available ( $n=86$ ). In patients with HER2-mutant disease, coincident mutations in TP53 and HER3 were enriched in patients with no clinical benefit (nominal  $P=0.018$  and  $P=0.064$ , respectively; Fig. 3). Although not significant after correcting for multiple hypothesis testing, potentially owing to the relatively small sample size, it is noteworthy that no patients with clinical benefit possessed co-mutation of HER2 and HER3. Concurrent mutation of these genes was observed

in multiple cancer types (breast  $n=3$ , bladder  $n=2$ , gastroesophageal  $n=2$ , colorectal  $n=1$  and pancreatic  $n=1$ ) and involved a variety of unique HER2 and HER3 mutations ( $n=8$  and  $n=9$ , respectively). Expanding our analysis to genomic activation at the pathway level, we identified somatic mutations of known oncogenic potential and grouped them by those involving the receptor tyrosine kinase (RTK)/RAS/RAF and PI3KCA/AKT/mTOR pathways, and cell cycle checkpoints (Extended Data Fig. 5b). In this analysis, concurrent aberrations in cell cycle checkpoints were associated with lack of clinical benefit ( $P=0.043$ ), and activation of RTK/RAS/RAF also trended towards a worse outcome ( $P=0.060$ ). The association between the cell-cycle pathway and lack of clinical benefit seems to be primarily driven by TP53 mutations, losing significance upon removal of TP53 mutations ( $P=0.769$ ). Interestingly, activation of the PI3K/AKT/mTOR pathway, an established negative predictor of response to HER2-targeted therapy in HER2-amplified breast cancer<sup>20–22</sup>, did not adversely affect the likelihood of clinical benefit ( $P=0.753$ ). It is possible that the clinical impact of concurrent gene/pathway activation may vary by tumour type, and future disease-specific studies are needed to define these associations better. Although these were exploratory analyses that will require confirmation, our results suggest that concurrent activation of specific genes as well as pathways may act as an additional modifier of response beyond cancer type and specific HER2 mutant allele.

## Discussion

The ability to profile cancer comprehensively at the point of care has made possible the opportunity to personalize therapy for each patient based on the compendium of genomic alterations identified<sup>23</sup>. Despite the promise of this approach, implementing this paradigm in clinical practice has been hampered by considerable gaps in knowledge about the biological and clinical importance of most genomic variants identified<sup>24</sup>. This challenge is exemplified by the marked diversity and wide distribution of HER2 and HER3 mutations in human cancers,



**Figure 3 | Genomic modifiers of response and outcome by treatment duration.** Comprehensive OncoPrint of the dichotomous clinical benefit groups for 86 patients with broad profiling data (left: no benefit ( $n = 66$ , biologically independent samples), right: clinical benefit ( $n = 20$ , biologically independent samples)). From top to bottom: TMB with the dotted line indicating the threshold for high TMB at 13.8 mutations (mut)

per megabase; microsatellite (MSI) status; allele/domain; tumour type; HER2 (*ERBB2*) status showing amplification; clonality and the presence of a single or multiple mutations; and co-alterations in genes associated with key pathways. \* $P = 0.064$ , \*\* $P = 0.018$ , Fisher's exact test. Statistical significance is lost when corrected for multiple hypothesis testing.

as well as by the difficulty of generating preclinical models of these mutations that correctly recreate their biology in patients. To our knowledge, SUMMIT provides the first comprehensive dataset on the clinical actionability of HER2 and HER3 mutations. We found that HER2 mutations are associated with HER2-dependence in a subset of patients with HER2-mutant tumours, but that response to HER kinase inhibition varies a function of the individual mutant variant, the tumour type as well as the pattern of co-mutations present.

Although we identified promising preliminary activity for neratinib in breast, biliary and cervical cancers, the response rate in these tumours was still lower than with approved therapies that target oncogenic alterations in *EGFR*, *ALK*, *ROS1* and *BRAF*. The low response rate in lung cancer, in which HER2 mutations exhibit mutually exclusivity with other known drivers<sup>25</sup>, is also notable and may in part reflect a lower potency of neratinib inhibition in Y772\_A775dup compared to other insertions or missense mutants<sup>18</sup>. Successfully targeting HER2 activation in other contexts has historically necessitated drug combinations. For example, single-agent trastuzumab has a response rate of only approximately 20% in *ERBB2*-amplified breast cancer<sup>26,27</sup>. By contrast, the overall survival in *ERBB2*-amplified breast and gastroesophageal cancers is markedly improved by adding trastuzumab to chemotherapy<sup>28,29</sup>. More recently, the intensification of HER2 inhibition through the combination of two HER2-targeted agents has been shown to result in synergistic efficacy in patients with *ERBB2*-amplified breast<sup>30–32</sup> or colorectal<sup>33,34</sup> cancers, as well as in HER2-mutant colorectal cancer xenografts<sup>6</sup>. Cumulatively, these data suggest that combining neratinib with another HER2-targeted therapy is a rational next step, and SUMMIT has been amended to evaluate this approach in multiple HER2-mutant tumour types.

SUMMIT represents a continued evolution in the design of basket studies, which enrol patients on the basis of qualifying mutations rather than tumour type. The initial generation of these studies focused on evaluating individual somatic mutations that were already clinically validated in one cancer (such as BRAF V600 in melanoma) in other tumour types<sup>10,35</sup>. More recently, basket studies have been used to generate initial or even practice-changing clinical data of truly novel genomic biomarkers, especially when these genomic alterations occur at low frequency across a wide distribution of cancer types<sup>11,36,37</sup>. SUMMIT extends this concept one step further by demonstrating for the first time how a single study can be used to simultaneously evaluate

a range of individual variants in HER2 and HER3, each with varying degrees of prior biologic characterization. This permissive enrolment strategy allowed us to treat patients harbouring mutations that, at the time of enrolment, had not been characterized preclinically as gain-of-function but were either recurrent or paralogous to known activating mutations in homologous genes. For example, patients with previously uncharacterized HER2 variants, such as V697L, D769N/H/Y and L869R, were included in this manner and responded to treatment, thus providing initial clinical proof-of-concept that these mutations confer a gain-of-function phenotype even before formal biologic characterization. The approach of pairing a permissive enrolment strategy with allele prioritization based on recurrence, paralogy and other readily computable features has potentially broad applicability to implementing genomic-driven oncology<sup>24</sup>. This strategy will take on even greater importance as clinical testing moves from targeted sequencing to whole exome or even whole genome sequencing, techniques that will allow for evaluation of an even greater number of therapeutic hypothesis but will also exponentially expand the number of uncharacterized alleles we routinely identify.

SUMMIT provides additional insights into the conduct of molecularly driven oncology studies. Our ability to understand the complex interactions between tumour lineage, individual HER2 variant and response to neratinib was only possible because of the relatively large size of this study ( $n = 141$ ). By comparison, many of the 'master/umbrella' protocols that are currently underway are designed to enrol a maximum of 30–40 patients into each genomically defined treatment arm. Our experience suggests that many studies of this size may be inadequately powered to identify the subgroups with true efficacy, assuming that most genomic alterations will not predict for tumour-type agnostic efficacy. SUMMIT also demonstrates the feasibility of enrolling patients based on local testing, with patients treated on the basis of 30 unique sequencing assays performed in 25 different laboratories. Despite this, concordance on retrospective central review was extremely high (96%).

An important impediment to progress in oncology has been the limited availability of preclinical model systems that accurately recreate the complex biology of human cancer. Although important strides have been made, the wide-scale profiling of cancer in the clinic provides the potentially transformative opportunity to interrogate cancer biology at the bedside in a manner previously only possible at the bench. Here,

we demonstrate how this opportunity can be leveraged to probe the biology of a diverse set of HER2 and HER3 mutations across a variety of solid tumours through pharmacological HER kinase inhibition in patients. In doing so, we found that response to pharmacological inhibition was based on the characteristics of both tumour type and genomic variant to a degree that was not predicted by established preclinical models. In summary, SUMMIT demonstrates how the clinical trial can become an important tool in refining our understanding of the biological dependencies in human cancers.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 4 August; accepted 22 December 2017.**

**Published online 31 January 2018.**

- Chmielecki, J. *et al.* Oncogenic alterations in *ERBB2/HER2* represent potential therapeutic targets across tumors from diverse anatomic sites of origin. *Oncologist* **20**, 7–12 (2015).
- Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
- Schram, A. *et al.* Landscape of somatic *ERBB2* mutations: Findings from AACR GENIE and comparison to ongoing *ERBB2* mutant basket study. *Cancer Res.* **77**, Abstract LB-103 (2017).
- Chang, M. T. *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34**, 155–163 (2016).
- Bose, R. *et al.* Activating *HER2* mutations in *HER2* gene amplification negative breast cancer. *Cancer Discov.* **3**, 224–237 (2013).
- Kavuri, S. M. *et al.* *HER2* activating mutations are targets for colorectal cancer treatment. *Cancer Discov.* **5**, 832–841 (2015).
- Jaiswal, B. S. *et al.* Oncogenic *ERBB3* mutations in human cancers. *Cancer Cell* **23**, 603–617 (2013).
- Chumsri, S. *et al.* Prolonged response to trastuzumab in a patient with *HER2*-nonamplified breast cancer with elevated *HER2* dimerization harboring an *ERBB2* S310F mutation. *J. Natl. Compr. Canc. Netw.* **13**, 1066–1070 (2015).
- Zabransky, D. J. *et al.* *HER2* missense mutations have distinct effects on oncogenic signaling and migration. *Proc. Natl Acad. Sci. USA* **112**, E6205–E6214 (2015).
- Hyman, D. M. *et al.* Vemurafenib in multiple nonmelanoma cancers with *BRAF* V600 mutations. *N. Engl. J. Med.* **373**, 726–736 (2015).
- Hyman, D. M. *et al.* AKT inhibition in solid tumors with AKT1 mutations. *J. Clin. Oncol.* **35**, 2251–2259 (2017).
- Ross, J. S. *et al.* A high frequency of activating extracellular domain *ERBB2* (*HER2*) mutation in micropapillary urothelial carcinoma. *Clin. Cancer Res.* **20**, 68–75 (2014).
- Ross, J. S. *et al.* Relapsed classic E-cadherin (*CDH1*)-mutated invasive lobular breast cancer shows a high frequency of *HER2* (*ERBB2*) gene mutations. *Clin. Cancer Res.* **19**, 2668–2676 (2013).
- Wolff, A. C. *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* **31**, 3997–4013 (2013).
- Ma, C. X. *et al.* Neratinib efficacy and circulating tumor DNA detection of *HER2* mutations in *HER2*-non-amplified metastatic breast cancer. *Clin. Cancer Res.* **23**, 5687–5695 (2017).
- Yasuda, H. *et al.* Structural, biochemical, and clinical characterization of epidermal growth factor receptor (*EGFR*) exon 20 insertion mutations in lung cancer. *Sci. Transl. Med.* **5**, 216ra177 (2013).
- Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).
- Kosaka, T. *et al.* Response heterogeneity of *EGFR* and *HER2* exon 20 insertions to covalent *EGFR* and *HER2* inhibitors. *Cancer Res.* **77**, 2712–2721 (2017).
- Freedman, R. A. *et al.* Translational Breast Cancer Research Consortium (TBCRC) 022: a phase II trial of neratinib for patients with human epidermal growth factor receptor 2-positive breast cancer and brain metastases. *J. Clin. Oncol.* **34**, 945–952 (2016).
- Shi, W. *et al.* Pathway level alterations rather than mutations in single genes predict response to *HER2*-targeted therapies in the neo-ALTTO trial. *Ann. Oncol.* **28**, 128–135 (2017).
- Loibl, S. *et al.* *PIK3CA* mutations are associated with reduced pathological complete response rates in primary *HER2*-positive breast cancer: pooled analysis of 967 patients from five prospective trials investigating lapatinib and trastuzumab. *Ann. Oncol.* **27**, 1519–1525 (2016).
- Baselga, J. *et al.* Biomarker analyses in CLEOPATRA: a phase III, placebo-controlled study of pertuzumab in human epidermal growth factor receptor 2-positive, first-line metastatic breast cancer. *J. Clin. Oncol.* **32**, 3753–3761 (2014).
- Schram, A. M., Berger, M. F. & Hyman, D. M. Precision oncology: charting a path forward to broader deployment of genomic profiling. *PLoS Med.* **14**, e1002242 (2017).
- Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing genome-driven oncology. *Cell* **168**, 584–599 (2017).
- Jordan, E. J. *et al.* Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discov.* **7**, 596–609 (2017).
- Baselga, J. *et al.* Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with *HER2*/neu-overexpressing metastatic breast cancer. *J. Clin. Oncol.* **14**, 737–744 (1996).
- Vogel, C. L. *et al.* Efficacy and safety of trastuzumab as a single agent in first-line treatment of *HER2*-overexpressing metastatic breast cancer. *J. Clin. Oncol.* **20**, 719–726 (2002).
- Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against *HER2* for metastatic breast cancer that overexpresses *HER2*. *N. Engl. J. Med.* **344**, 783–792 (2001).
- Bang, Y. J. *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of *HER2*-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376**, 687–697 (2010).
- Swain, S. M. *et al.* Pertuzumab, trastuzumab, and docetaxel in *HER2*-positive metastatic breast cancer. *N. Engl. J. Med.* **372**, 724–734 (2015).
- Blackwell, K. L. *et al.* Randomized study of lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *J. Clin. Oncol.* **28**, 1124–1130 (2010).
- Baselga, J. *et al.* Lapatinib with trastuzumab for *HER2*-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet* **379**, 633–640 (2012).
- Bertotti, A. *et al.* A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies *HER2* as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov.* **1**, 508–523 (2011).
- Sartore-Bianchi, A. *et al.* Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, *KRAS* codon 12/13 wild-type, *HER2*-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17**, 738–746 (2016).
- Kaufman, B. *et al.* Olaparib monotherapy in patients with advanced cancer and a germline *BRCA1/2* mutation. *J. Clin. Oncol.* **33**, 244–250 (2015).
- Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
- Hyman, D. M. *et al.* The efficacy of larotrectinib (LOXO-101), a selective tropomyosin receptor kinase (TRK) inhibitor, in adult and pediatric *TRK* fusion cancers. *J. Clin. Oncol.* **35**, abstract LBA2501 (2017).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank patients and their families for participating in this study. Editorial support, not including writing, was provided by L. Miller. This work was funded by Puma Biotechnology, and supported by grants from the National Institutes of Health (grants P30 CA008748, P30 CA016672, P30 CA014089, R01 CA204749, R01 CA80195, T32 CA009207, 1U01 CA180964 and UL1 TR000371), the National Institutes of Health/National Cancer Institute (Breast SPORE grant P50 CA098131), Cycle for Survival, Marie-Josée and Henry R. Kravis Center for Molecular Oncology, The Cancer Prevention and Research Institute of Texas (RP1100584), the Sheikh Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, Nellie B. Connally Breast Cancer Research Endowment, and the Breast Cancer Research Foundation.

**Author Contributions** D.M.H., H.W., M.F.B., R.E.C., F.X., A.B., L.D.E., G.M., C.F., A.S.L., R.P.B., J.B. and D.B.S. designed the study and supervised the analyses. R.E.C., F.X., L.D.E., G.M., C.F., A.S.L. and R.P.B. helped to collect and monitor the clinical outcome data. D.M.H., S.A.P., J.R., C.S., G.I.S., D.J., D.I.Q., V.M., B.D., I.A.M., V.B., E.C., S.L., A.C.L., J.P.E., B.T.L., A.J.H., R.M., A.M.S., A.D., L.M.S., K.J., G.I., J.J.H., C.L.A., F.M.B., J.B. and D.B.D. enrolled patients and provided patient samples. G.U. developed the PET response criteria and performed radiographic response assessments. B.S.T., J.P., J.T., S.D.S., N.B., M.M., M.F.B., J.B. and D.B.S. performed the tumour and plasma sequencing, provided computational infrastructure, and made final variant calls. D.M.H., H.W., M.S., B.S.T., J.P., J.T., H.B., M.F.B. and D.B.S. analysed clinical and genomic data and performed the integrated efficacy analyses. F.X. performed biostatistical analyses of the clinical efficacy data. D.M.H., H.W., B.S.T., C.L.A., F.M.B. and D.B.S. wrote the manuscript with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.M.H. (hymand@mskcc.org).

**Reviewer Information** Nature thanks E. Mardis and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Patients.** Eligible patients had histologically confirmed advanced solid tumours harbouring HER2 or HER3 mutations, an Eastern Cooperative Oncology Group (ECOG) performance score of 0–2 and an unlimited number of previous therapies. Patients with previous exposure to HER kinase inhibitors and unstable brain metastases were excluded. HER2 and HER3 mutations were determined by local tumour testing as routinely performed or ordered by each participating site. In total, 85% (120 out of 141) of enrolled patients were identified by next-generation sequencing assays. In 81% of cases (97 out of 120), the next-generation sequencing assay included full exon coverage for *ERBB2* or *ERBB3*, whereas in 19% (23 out of 120) of cases, only select exons or hotspots were included in the assay design. The remaining 15% (21 out of 141) of patients were enrolled via RT-PCR, Sanger, pyrosequencing, or mass spectrometry-based sequencing methods. The study was approved by the institutional review board or independent ethics committee at each site and complied with the International Ethical Guidelines for Biomedical Research Involving Human Subjects, Good Clinical Practice guidelines, the Declaration of Helsinki, and local laws. Written informed consent was obtained from all participants.

**Study design, treatment and endpoints.** This was a multi-cohort basket study of patients with solid tumours harbouring HER2 and HER3 mutations. Patients with HER2-mutant tumours were enrolled into one of several disease-specific cohorts or an 'other' cohort for tumour types not otherwise specified; all patients with HER3-mutant tumours were enrolled to one cohort. Patients known to contain both HER2 and HER3 mutations at the time of enrolment were assigned to the HER2-mutant cohort. Patients were treated with neratinib 240 mg daily on a continuous basis with mandatory loperamide prophylaxis during cycle 1. The primary endpoint was ORR<sub>s</sub>, as assessed by investigators according to RECIST (version 1.1). Secondary endpoints included best overall response, progression-free survival, overall survival and safety. Patients who were not evaluable by RECIST were permitted to enrol and were evaluated for response by <sup>18</sup>F-fluorodeoxyglucose PET according to a modified version of the original PET Response Criteria in Solid Tumours (PERCIST; version 1.0)<sup>38</sup>, referred to here as PET Response Criteria (PRC, Extended Data Table 3).

**Assessments.** Disease assessments with computed tomography, magnetic resonance imaging or combined positron emission tomography-computed tomography (for those evaluated by PRC) were performed at baseline and then every 8 weeks until disease progression, death or withdrawal. Adverse events were graded by the investigator according to the Common Terminology Criteria for Adverse Events (version 4.0) until day 28 after discontinuation of study treatment.

**Genomic biomarker studies.** All samples were assigned anonymized identifiers by the study sponsor based on the order of study enrolment. Both tumour DNA and tumour-derived cell-free DNA in plasma were collected with the goals of confirming locally reported HER2/3 mutations as well as evaluating how *ERBB2* and *ERBB3* copy number and clonality as well as co-mutational pattern affected outcome. Collection of archival tumour and plasma samples was mandatory for all patients. Next-generation sequencing was performed using targeted sequencing of pretreatment DNA from formalin-fixed paraffin-embedded tumour and matched blood specimens (preferentially) and cell-free DNA (if tumour was not available or was inadequate). A custom single-gene *ERBB2* capture next-generation sequencing test was also performed on pretreatment cell-free DNA in a subset of patients with HER2-mutant disease.

**Central sequencing confirmation.** For patients with adequate material, DNA from formalin-fixed paraffin-embedded ( $n=91$ ) or tumour-derived cell-free DNA from plasma ( $n=15$ ) and matched germline DNA ( $n=102$ ) underwent targeted next-generation sequencing assay using Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT)<sup>2</sup>, producing an average of 738-fold coverage per tumour (range: 253–1,383). In brief, this assay uses a hybridization-based exon capture designed to capture all protein-coding exons and select introns of oncogenes, tumour-suppressor genes and key members of pathways that may be actionable by targeted therapies. In this study, either 341 ( $n=18$ ) or 410 ( $n=88$ ) key cancer-associated genes were analysed (Supplementary Information). Sequencing data were analysed as previously described to identify somatic single-nucleotide variants, small insertions and deletions, copy number alterations and structural arrangements<sup>39</sup>. In addition, hotspot alterations were identified using an adaptation of a previously described method<sup>4</sup> applied to a cohort of 24,592 sequenced human cancers<sup>40</sup>. For gene-level analysis, select genes within our targeted 341/410 MSK-IMPACT panel involved in the RTK/RAS/RAF, PIK3CA/AKT/mTOR, and cell cycle checkpoint pathways were selected using the KEGG pathway database<sup>41</sup>. For pathway level analysis, only potentially oncogenic alterations in the selected genes were included and determined to be oncogenic by OncoKB (version September 2017), a curated knowledge base of the oncogenic effects and treatment implications of mutations and cancer genes (<http://www.oncokb.org>)<sup>42</sup>.

**HER2 amplification and clonality analysis.** For patients in the HER2-mutant arm with MSK-IMPACT sequencing data (with matched germline DNA,  $n=74$ ), the Fraction and Allele-Specific Copy Number Estimates from Tumour Sequencing (FACETS) algorithm (version 0.3.9) was used to estimate tumour purity and ploidy,

and total and allele-specific copy number<sup>43</sup>. Tumour samples with purity less than 20% were excluded from this analysis. Focal HER2 amplifications for tumours with MSK-IMPACT and FACETS data were inferred using the following criteria: fold change  $\geq 1.5$  (MSK-IMPACT tumour:normal sequencing coverage ratio) and total HER2 copy number  $\geq 4$  copies (FACETS-derived total copy number). To infer clonality of each HER2 mutation, cancer cell fractions were estimated with 95% confidence intervals by integrating FACETS-derived joint segmentation and MSK-IMPACT mutation data as input into the ABSOLUTE algorithm<sup>44</sup> (version 1.0.6). Mutations were classified as either clonal or subclonal based on the following criteria: clonal if the estimated cancer cell fractions  $> 0.85$ , otherwise subclonal. For patients with HER2 amplification, the mutation copy number (mutation multiplicity) was calculated as previously described<sup>45</sup> to infer amplification of the mutant allele when the mutation multiplicity was greater than half of the total HER2 copy number.

**TMB and MSI.** TMB, defined as the number of non-synonymous mutations per megabase, was calculated for patients with MSK-IMPACT sequencing data ( $n=106$ )<sup>6</sup>. MSI was assessed for patients with HER2-mutant tumours with matched germline DNA sequencing data ( $n=89$ ) using an orthogonal bioinformatics tool, MSIsensor<sup>46</sup>. Furthermore, mutations were decomposed into the 30 constituent mutational signatures as described previously<sup>47</sup>. In brief, MSIsensor scores  $< 10$  were classified as microsatellite stable and  $> 10$  were considered MSI-high using a previously validated cut-off score<sup>48</sup>. Those with a MSIsensor score of  $< 10$  but having evidence of a dominant mismatch repair mutational signature were also considered MSI<sup>43,47</sup>.

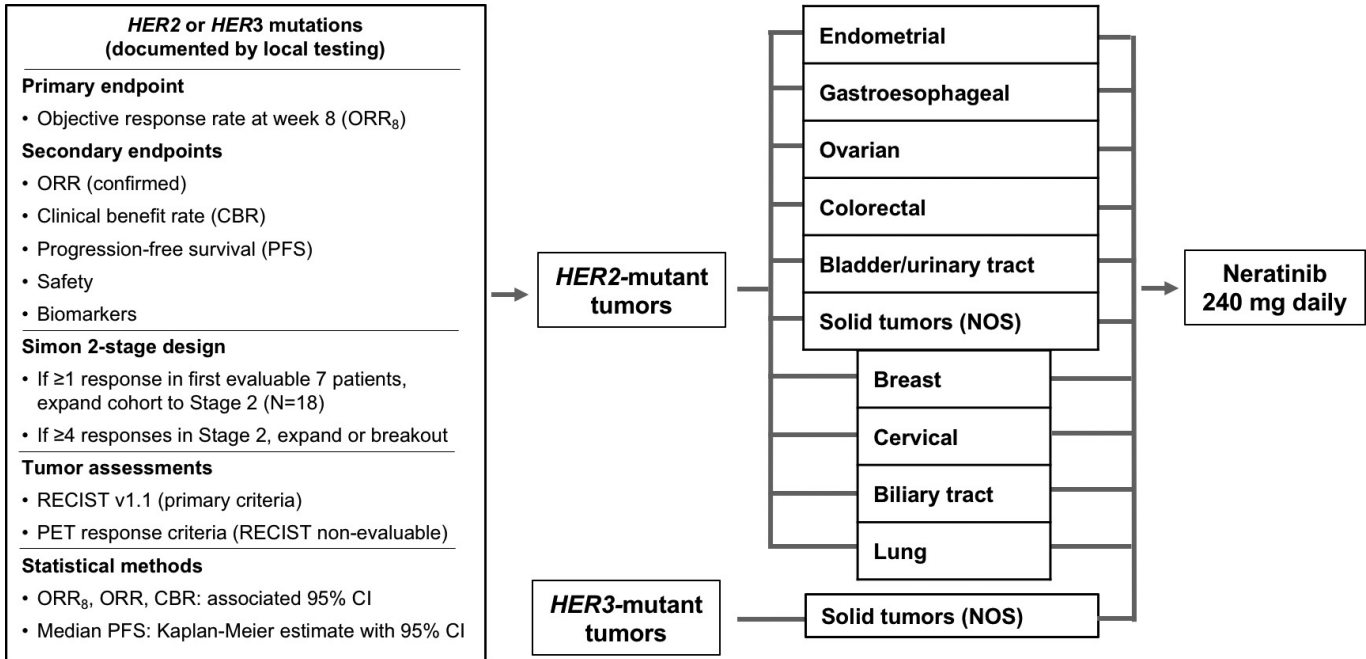
**Statistical analysis.** For each HER2-mutant tumour type and the HER3-mutant cohort, a Simon optimal two-stage design with a true ORR<sub>s</sub>  $\leq 10\%$  was considered unacceptable (null hypothesis), whereas a true ORR<sub>s</sub>  $\geq 30\%$  (alternative hypothesis) merited further study. Efficacy in each cohort was analysed independently and the study was not designed to compare efficacy across cohorts formally. All patients who received at least one dose of neratinib were included in the safety and efficacy cohorts. All data reflect an interim data-cut taken on 10 March 2017 from patients enrolled up to 16 December 2016 (Extended Data Fig. 6). Most patients were off therapy at the time of data analysis (Extended Data Table 4). Progression-free survival was estimated using the Kaplan–Meier method. The study is registered at <http://www.clinicaltrials.gov>, under the identifier NCT01953926. Individual associations among genomic changes and response were assessed by either Fisher's exact or chi-squared tests (where appropriate) and corrected for multiple hypothesis testing using Benjamini–Hochberg correction.

Chi-squared or Fisher's exact tests were performed to compare gene-level and pathway-level associations between the dichotomous clinical benefit groups. *P* values were corrected for multiple hypothesis testing using Benjamini–Hochberg correction. HER2 and HER3 lollipop distribution plots were generated using ProteinPaint<sup>49</sup>. All other figures were generated using R software (<http://www.R-project.org/>).

This clinical trial was not randomized and investigators were not blinded to treatment allocation and outcome assessment.

**Data availability.** All datasets generated during and/or analysed during the current study, including patient-level clinical data as well as all sequencing data have been deposited and are publicly available in the cBioPortal for Cancer Genomics under the accession code 'SUMMIT, Nature, 2018' ([http://www.cbioportal.org/study?id=summit\\_2018](http://www.cbioportal.org/study?id=summit_2018)).

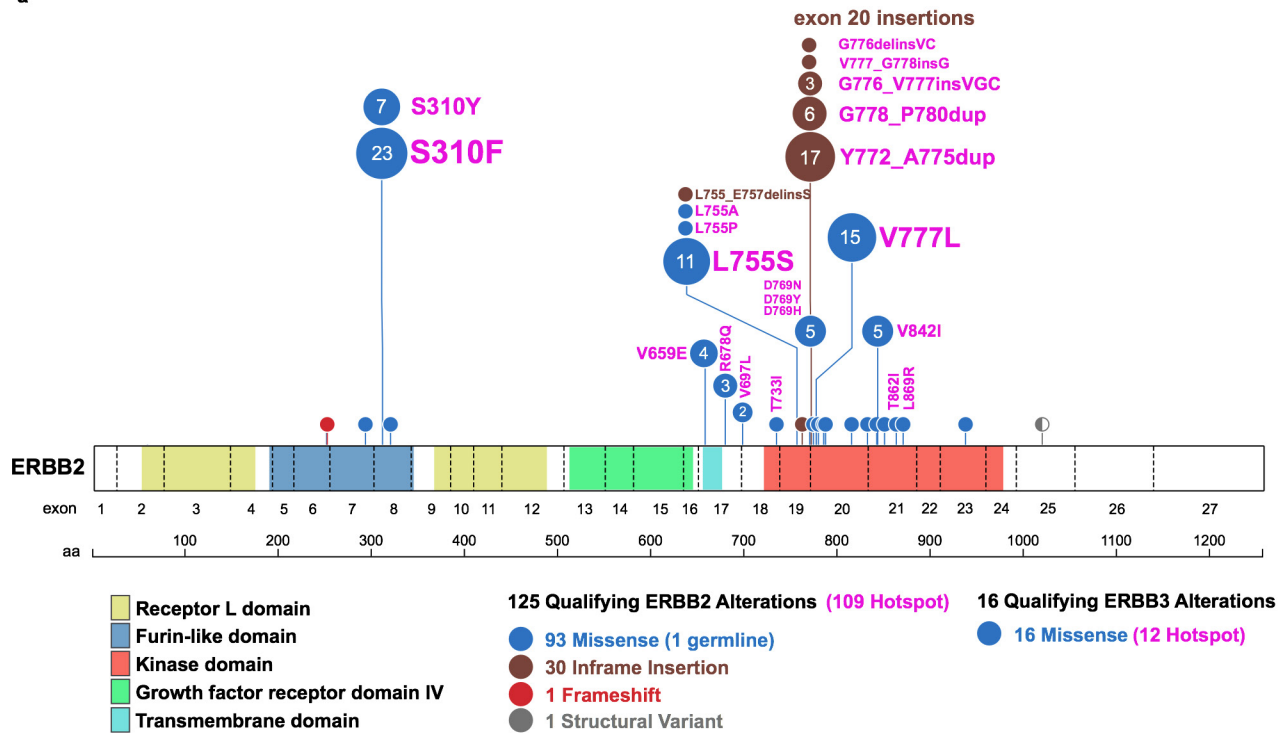
38. Wahl, R. L., Jacene, H., Kasamon, Y. & Lodge, M. A. From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. *J. Nucl. Med.* **50**, 122S–150S (2009).
39. Cheng, D. T. et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).
40. Chang, M. T. et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* <http://doi.org/10.1158/2159-8290.CD-17-0321> (2017).
41. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
42. Chakravarty D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* <http://doi.org/10.1200/PO.17.00011> (2017).
43. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
44. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
45. McGranahan, N. et al. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* **7**, 283ra54 (2015).
46. Niu, B. et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016 (2014).
47. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
48. Middha et al. Reliable pan-cancer microsatellite instability assessment by using targeted next-generation sequencing data. *JCO Precis. Oncol.* <http://doi.org/10.1200/PO.17.00084> (2017).
49. Zhou, X. et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.* **48**, 4–6 (2016).



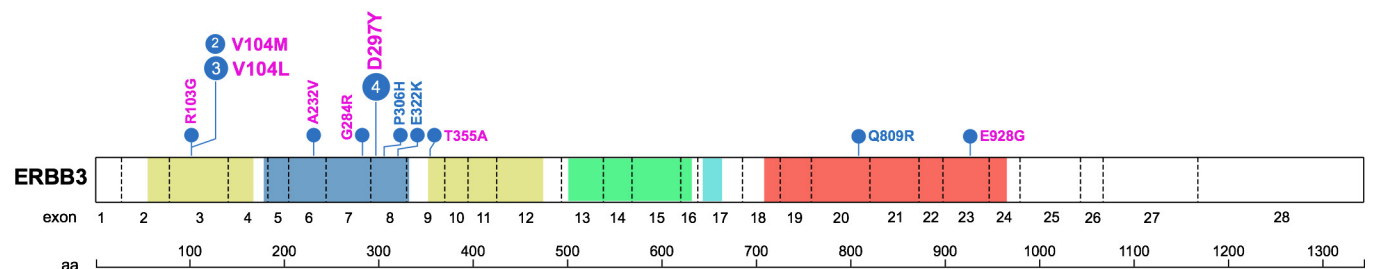
**Extended Data Figure 1 | Design of SUMMIT study.** Five tumour-specific HER2 (*ERBB2*)-mutant cohorts were pre-specified (endometrial, gastroesophageal, ovarian, colorectal and bladder/urinary tract). In addition, a sixth ‘solid tumour (not otherwise specified, NOS)’ HER2-mutant cohort allowed for the enrolment of patients with any other cancer types. A sufficient number of patients with breast, cervical, biliary and

lung cancer were enrolled in the solid tumours (NOS) cohort to permit independent efficacy analysis using the same design as the pre-specified cohorts. Patients with HER3 (*ERBB3*)-mutant tumours were enrolled in a HER3-specific cohort regardless of tumour type. CBR, clinical benefit rate; cfDNA, cell-free (tumour) DNA; CI, confidence interval.

a

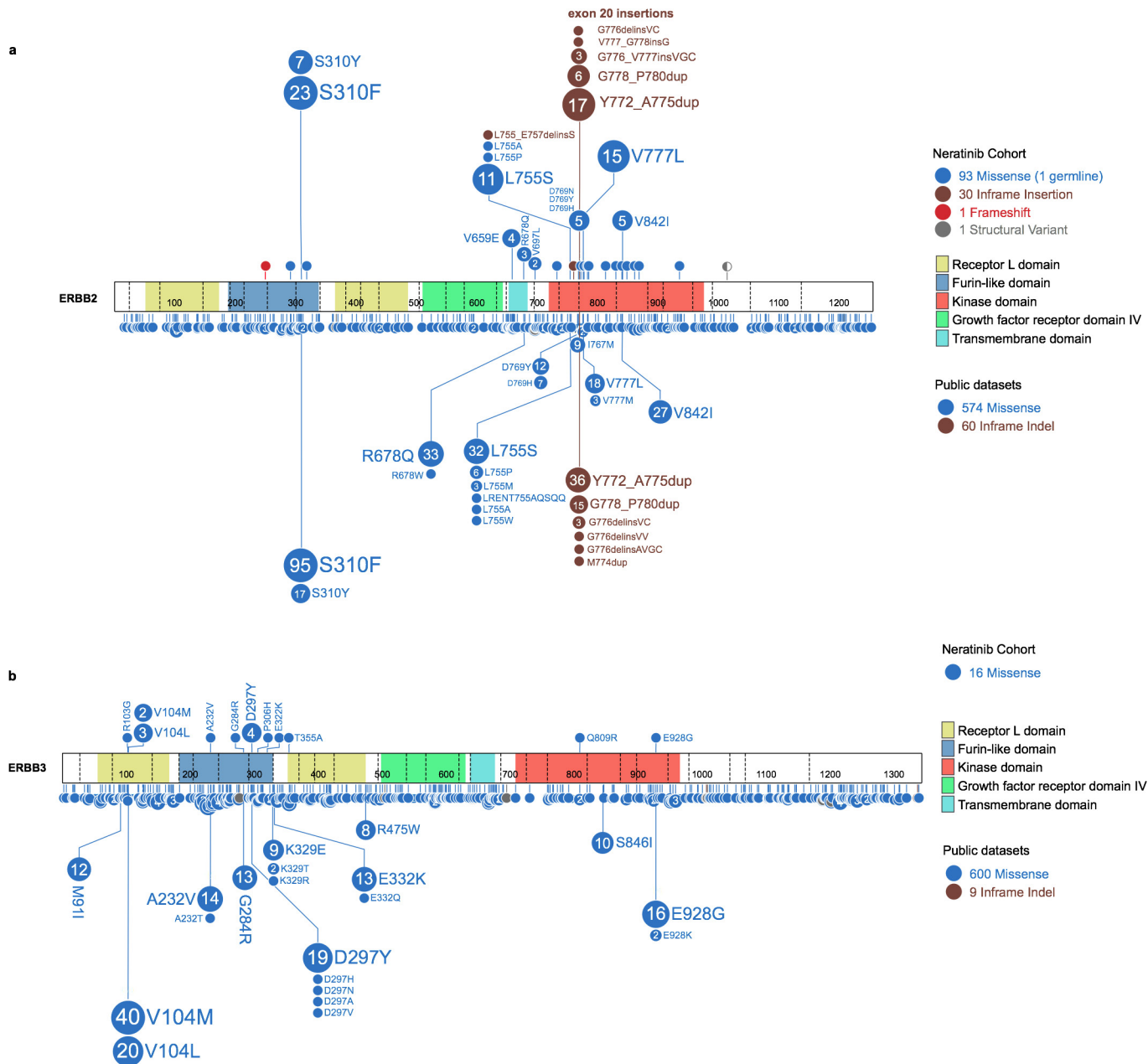


b



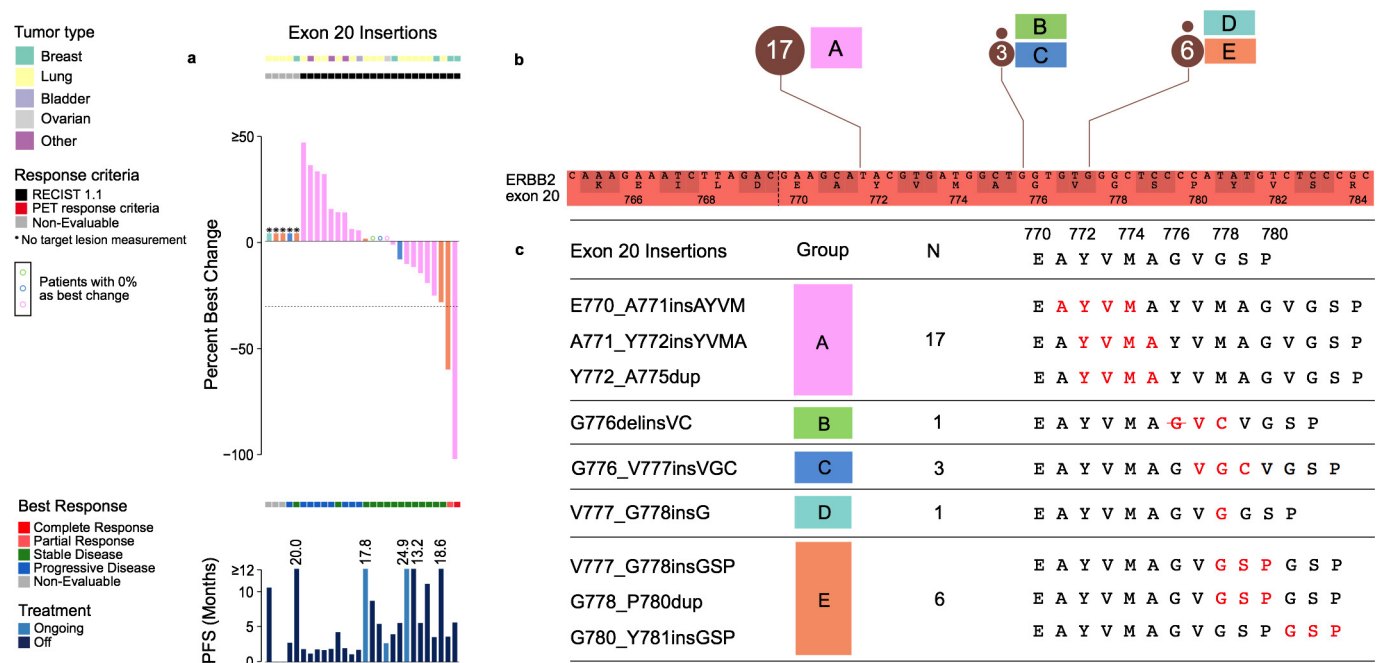
**Extended Data Figure 2 | Distribution of HER2 and HER3 mutations positioned by their amino acid coordinates across the respective protein domains. a, b, HER2 (a) and HER3 (b) mutations (125 and 16 mutations, respectively). Each unique mutation is represented by a circle,**

**with the circle size and number representing the frequency, and coloured to show the mutation class as indicated in the legend. The corresponding amino acid change and common hotspot mutations (shown in pink) are labelled next to the circles.**



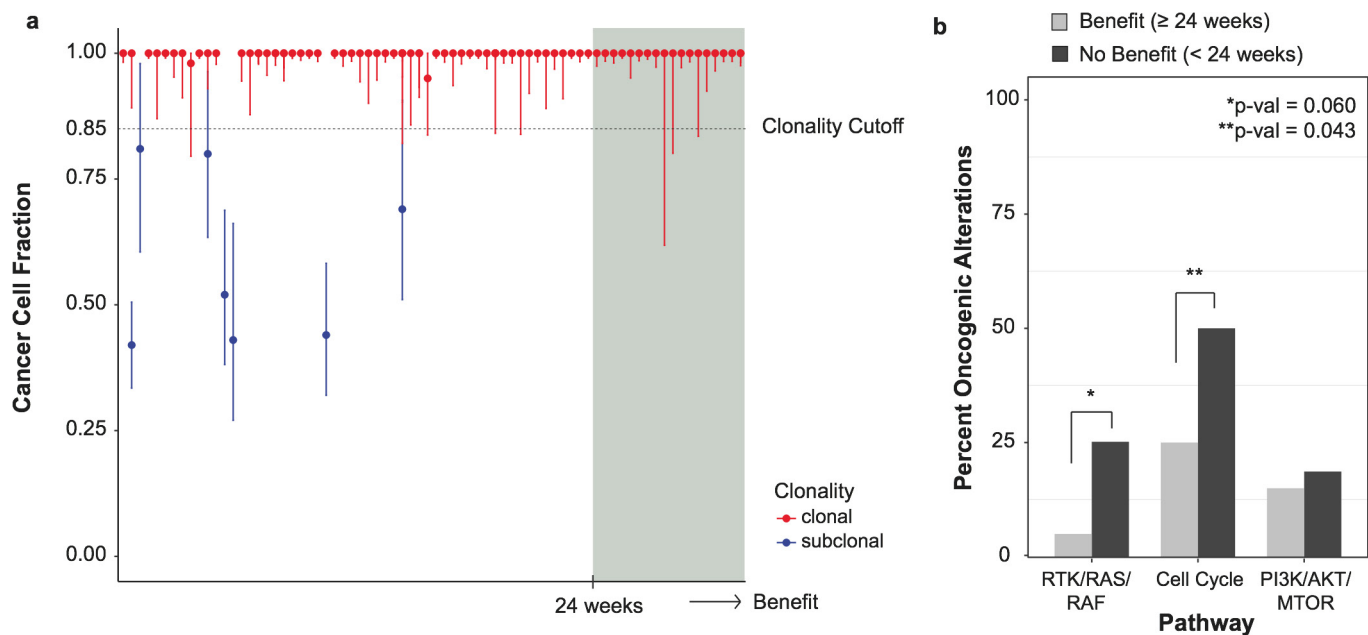
**Extended Data Figure 3 | Spectrum of HER2 and HER3 mutations observed in the neratinib study versus TCGA, ICGC and other public datasets. a, b, Distribution of HER2 (a) and HER3 (b) mutations observed**

across our cohort in comparison to the spectrum of HER2 and HER3 mutations (reflected lollipop) from publically available datasets (TCGA, ICGC and other published studies).



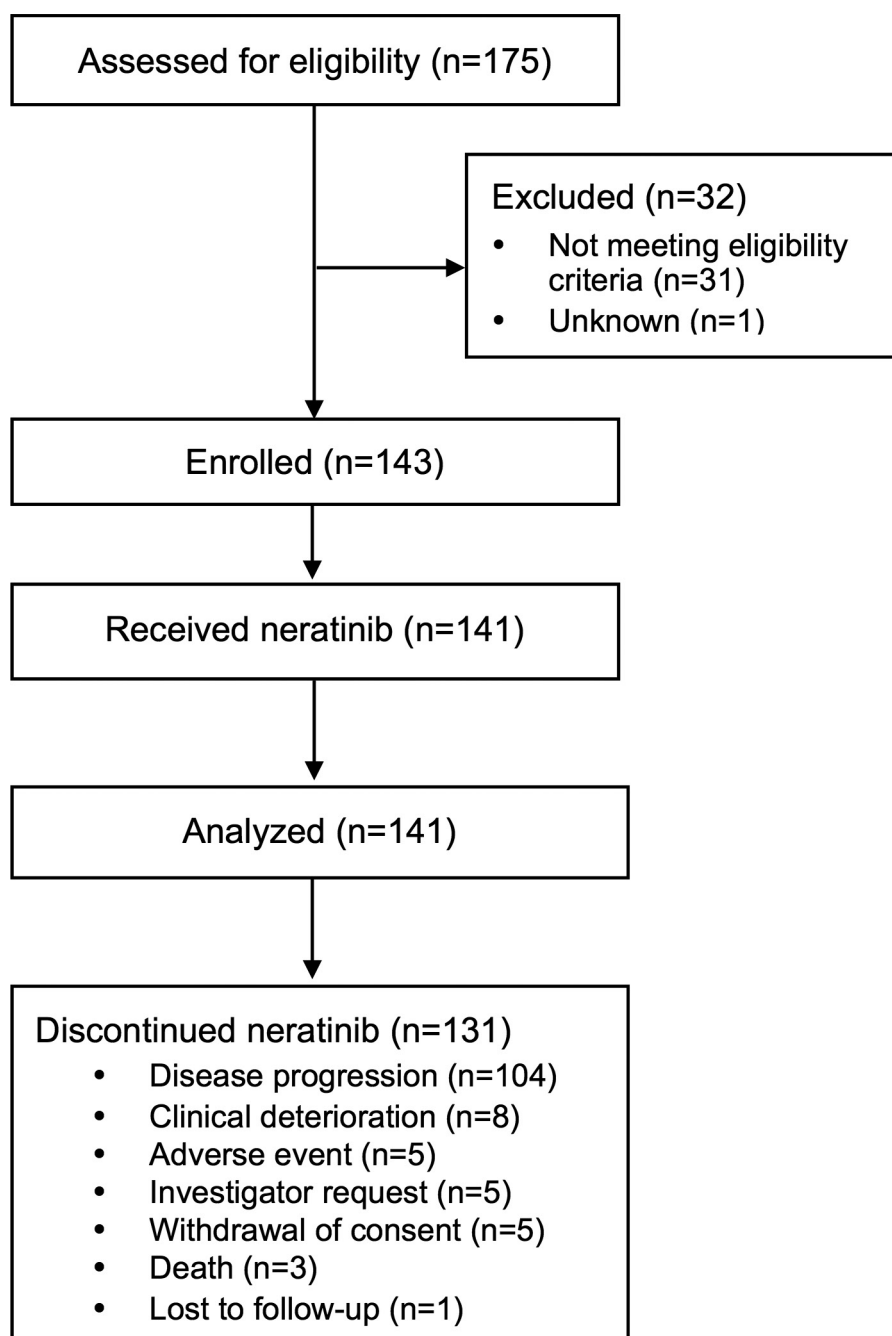
**Extended Data Figure 4 | Distribution and outcome of 28 HER2 exon 20 insertions.** **a**, Percentage best change and PFS plots corresponding to each type of exon 20 insertion (colour coded by synonymous amino acid change). Three cases with no change are indicated in colour-coded

circles above the x axis. **b**, Zoomed-in schematic of all exon 20 insertions positioned by their amino acid coordinates and frequencies. **c**, Five unique types of exon 20 insertions observed in the study with the resulting full amino acid sequences (insertion indicated in red).



**Extended Data Figure 5 | Genomic modifiers of response and outcome by treatment duration.** **a**, Cancer cell fractions with 95% confidence intervals and clonality status of all HER2 mutations in 74 patients with sufficient sequencing data ordered by increasing clinical benefit (weeks on

therapy). **b**, Comparison of the percentage activation of known oncogenic alterations in the three pathways between the patients of clinical benefit ( $n = 20$ , biologically independent samples) and no benefit ( $n = 66$ , biologically independent samples). Nominal Fisher's  $P$  values are shown.



Extended Data Figure 6 | SUMMIT CONSORT diagram.

Extended Data Table 1 | Patient demographics and efficacy by cohort

| Characteristic  | HER2               |                   |                   |                      |                        |                   |                      | HER3                           |                   |                   |                   |
|---|--------------------|-------------------|-------------------|----------------------|------------------------|-------------------|----------------------|--------------------------------|-------------------|-------------------|-------------------|
|   | Breast<br>(n=25)   | Lung<br>(n=26)    | Bladder<br>(n=16) | Colorectal<br>(n=12) | Biliary tract<br>(n=9) | Cervical<br>(n=5) | Endometrial<br>(n=7) | Gastro-<br>esophageal<br>(n=5) | Ovarian<br>(n=4)  | NOS<br>(n=16)     | NOS<br>(n=16)     |
| Median (range), years                                   | 57.0 (37-80)       | 62.0 (46-74)      | 65.0 (48-83)      | 65.0 (30-81)         | 66.0 (57-78)           | 49.0 (42-56)      | 57.0 (54-74)         | 67.0 (36-70)                   | 56.5 (38-58)      | 59.0 (32-80)      | 66.0 (39-82)      |
| <65 years, n (%)  | 19 (76.0)          | 18 (69.2)         | 8 (50.0)          | 6 (50.0)             | 2 (22.2)               | 5 (100)           | (71.4)               | 1 (20.0)                       | 4 (100)           | 13 (81.3)         | 7 (43.8)          |
| ≥65 years, n (%)  | 6 (24.0)           | 8 (30.8)          | 8 (50.0)          | 6 (50.0)             | 7 (77.8)               | 0 (0)             | 2 (28.6)             | 4 (80.0)                       | 0 (0)             | 3 (18.8)          | 9 (56.3)          |
| Sex, n (%)  |                    |                   |                   |                      |                        |                   |                      |                                |                   |                   |                   |
| Female  | 24 (96.0)          | 17 (65.4)         | 3 (18.8)          | 6 (50.0)             | 5 (55.6)               | 5 (100)           | 7 (100)              | 2 (40.0)                       | 4 (100)           | 7 (43.8)          | 12 (75.0)         |
| Male  | 1 (4.0)            | 9 (34.6)          | 13 (81.3)         | 6 (50.0)             | 4 (44.4)               | 0 (0)             | 0 (0)                | 3 (60.0)                       | 0 (0)             | 9 (56.3)          | 4 (25.0)          |
| ECOG PS, n (%)  |                    |                   |                   |                      |                        |                   |                      |                                |                   |                   |                   |
| 0   | 7 (28.0)           | 11 (42.3)         | 6 (37.5)          | 5 (41.7)             | 2 (22.2)               | 1 (20.0)          | 2 (28.6)             | 0 (0)                          | 0 (0)             | 3 (18.8)          | 1 (6.3)           |
| 1   | 17 (68.0)          | 14 (53.8)         | 10 (62.5)         | 7 (58.3)             | 6 (66.7)               | 4 (80.0)          | 5 (71.4)             | 5 (100)                        | 4 (100)           | 11 (68.8)         | 12 (75.0)         |
| 2   | 1 (4.0)            | 1 (3.8)           | 0 (0)             | 0 (0)                | 1 (11.1)               | 0 (0)             | 0 (0)                | 0 (0)                          | 0 (0)             | 2 (12.5)          | 3 (18.8)          |
| Prior systemic lines, n (%)                             |                    |                   |                   |                      |                        |                   |                      |                                |                   |                   |                   |
| None  | 0 (0)              | 1 (3.8)           | 1 (6.3)           | 0 (0)                | 1 (11.1)               | 0 (0)             | 0 (0)                | 0 (0)                          | 0 (0)             | 1 (6.3)           | 0 (0)             |
| 1   | 3 (12.0)           | 12 (46.2)         | 2 (12.5)          | 4 (33.3)             | 3 (33.3)               | 0 (0)             | 1 (14.3)             | 2 (40.0)                       | 0 (0)             | 6 (37.5)          | 1 (6.3)           |
| 2   | 2 (8.0)            | 6 (23.1)          | 9 (56.3)          | 3 (25.0)             | 2 (22.2)               | 3 (60.0)          | 2 (28.6)             | 1 (20.0)                       | 0 (0)             | 2 (12.5)          | 11 (68.8)         |
| ≥3  | 20 (80.0)          | 7 (26.9)          | 4 (25.0)          | 5 (41.7)             | 3 (33.3)               | 2 (40.0)          | 4 (57.1)             | 2 (40.0)                       | 4 (100)           | 7 (43.8)          | 4 (25.0)          |
| Median time from metastasis to enrolment, years (range) | 2.64<br>(0.1-15.0) | 0.83<br>(0.1-3.1) | 0.69<br>(0.2-2.3) | 1.14<br>(0.0-2.7)    | 1.00<br>(0.0-2.8)      | 1.40<br>(0.3-4.5) | 0.43<br>(0.2-4.4)    | 0.80<br>(0.4-4.3)              | 7.54<br>(1.1-7.7) | 1.35<br>(0.0-5.4) | 1.13<br>(0.3-4.5) |
| Outcome   | HER2               |                   |                   |                      |                        |                   |                      | HER3                           |                   |                   |                   |
|   | Breast<br>(n=25)   | Lung<br>(n=26)    | Bladder<br>(n=16) | Colorectal<br>(n=12) | Biliary tract<br>(n=9) | Cervical<br>(n=5) | Endometrial<br>(n=7) | Gastroeso-<br>phageal (n=5)    | Ovarian<br>(n=4)  | NOS<br>(n=16)     | NOS<br>(n=16)     |
| ORR at week 8, n (%)                                    | 8 (32.0)           | 1 (3.8)           | 0 (0.0)           | 0 (0.0)              | 2 (22.2)               | 1 (20.0)          | 0 (0.0)              | 0 (0.0)                        | 0 (0.0)           | 1 (6.3)           | 0 (0.0)           |
| [95% CI]  | [14.9–53.5]        | [0.1–19.6]        | [0.0–20.6]        | [0.0–26.5]           | [2.8–60.0]             | [0.5–71.6]        | [0.0–41.0]           | [0.0–52.2]                     | [0.0–60.2]        | [0.2–30.2]        | [0.0–20.6]        |
| ORR, n (%)  | 6 (24.0)           | 1 (3.8)           | 0 (0.0)           | 0 (0.0)              | 0 (0.0)                | 1 (20.0)          | 0 (0.0)              | 0 (0.0)                        | 0 (0.0)           | 0 (0.0)           | 0 (0.0)           |
| [95% CI]  | [9.4–45.1]         | [0.1–19.6]        | [0.0–20.6]        | [0.0–26.5]           | [0.0–33.6]             | [0.5–71.6]        | [0.0–41.0]           | [0.0–52.2]                     | [0.0–60.2]        | [0.0–20.6]        | [0.0–20.6]        |
| Clinical benefit rate, n (%)                            | 10 (40.0)          | 11 (42.3)         | 3 (18.8)          | 1 (8.3)              | 3 (33.3)               | 3 (60.0)          | 2 (28.6)             | 1 (20.0)                       | 0 (0.0)           | 3 (18.8)          | 2 (12.5)          |
| [95% CI]  | [21.1–61.3]        | [23.4–63.1]       | [4.0–45.6]        | [0.2–38.5]           | [7.5–70.1]             | [14.7–94.7]       | [3.7–71.0]           | [0.5–71.6]                     | [0.0–60.2]        | [4.0–45.6]        | [1.6–38.3]        |
| Median PFS, months                                      | 3.5                | 5.5               | 1.8               | 1.8                  | 2.8                    | 20.1              | 2.6                  | 1.7                            | 2.1               | 1.9               | 1.7               |

Extended Data Table 2 | Treatment-emergent adverse events (occurring in  $\geq 10\%$  of patients)

| Adverse event, n (%)                 | Neratinib monotherapy (N=141) |                        |
|--------------------------------------|-------------------------------|------------------------|
|                                      | Any grade                     | Grade $\geq 3$         |
| Diarrhoea                            | 104 (73.8)                    | 31 (22.0) <sup>*</sup> |
| Nausea                               | 61 (43.3)                     | 3 (2.1)                |
| Vomiting                             | 58 (41.1)                     | 3 (2.1)                |
| Constipation                         | 49 (34.8)                     | 2 (1.4)                |
| Fatigue                              | 45 (31.9)                     | 5 (3.5)                |
| Decreased appetite                   | 40 (28.4)                     | 1 (0.7)                |
| Abdominal pain                       | 33 (23.4)                     | 7 (5.0)                |
| Anaemia                              | 22 (15.6)                     | 10 (7.1)               |
| Dyspnoea                             | 18 (12.8)                     | 5 (3.5)                |
| Dehydration                          | 17 (12.1)                     | 8 (5.7)                |
| Aspartate aminotransferase increased | 15 (10.6)                     | 5 (3.5)                |
| Asthenia                             | 15 (10.6)                     | 1 (0.7)                |
| Weight decreased                     | 15 (10.6)                     | 0                      |

**Characteristics of diarrhoea**

|   |           |
|---|-----------|
| Action taken with neratinib, n (%)                              |           |
| Permanent discontinuation                                       | 4 (2.8)   |
| Serious <sup>†</sup> diarrhoea, n (%)                           | 15 (10.6) |
| Median (range) number of grade 3 diarrhoea episodes per patient | 1 (1–12)  |
| Median (range) duration of grade 3 diarrhoea episode, days      | 2 (1–8)   |
| Median (range) time to first grade 3 diarrhoea episode, days    | 10 (4–87) |

\*All events of grade 3.

†Serious adverse event as defined per study protocol.

**Extended Data Table 3 | PET response criteria**

| Response category                   | Based on sum of $SUV_{max}$ from 1 to 5 target lesions. Each target lesion with initial $SUV_{max}$ of $>1.5 \times$ normal liver background $SUV_{max}$  |
|-------------------------------------|---|
| Complete metabolic response (CMR)   | <ul style="list-style-type: none"> <li>Reduction of <math>SUV_{max}</math> of all target lesions to less than normal liver background <math>SUV_{max}</math> (for non-brain lesions) or less than normal brain background <math>SUV_{max}</math> (for brain lesions)</li> </ul> AND <ul style="list-style-type: none"> <li>The reduction of all other FDG-avid lesions consistent with disease to less than normal liver background <math>SUV_{max}</math></li> </ul> |
| Partial metabolic response (PMR)    | <ul style="list-style-type: none"> <li>Sum of <math>SUV_{max}</math> of all target lesions is decreased by <math>\geq 30\%</math> compared to baseline sum of <math>SUV_{max}</math> of all target lesions</li> </ul> AND <ul style="list-style-type: none"> <li>No new lesions</li> </ul>  |
| Stable metabolic disease (SMD)      | Not satisfying the criteria for CMR, PMR, PMD, or NE  |
| Progressive metabolic disease (PMD) | <ul style="list-style-type: none"> <li>Sum of <math>SUV_{max}</math> of all target lesions is increased by <math>\geq 30\%</math></li> </ul> OR <ul style="list-style-type: none"> <li>Appearance of one or more unequivocal new FDG-avid lesions</li> </ul>  |
| Not evaluable (NE)                  | <ul style="list-style-type: none"> <li>Missing FDG-PET series or incomplete anatomy at follow-up timepoint</li> <li>A PET/CT scanner change from baseline</li> <li>Variation in FDG uptake time <math>\geq 15</math> minutes compared to baseline</li> <li>Change in reconstruction algorithm</li> </ul>  |

CT, computed tomography; FDG-PET,  $^{18}\text{F}$ -fluorodeoxyglucose positron-emission tomography;  $SUV_{max}$ , maximum standardized uptake value.

Extended Data Table 4 | Patient disposition by cohort

| Characteristic                             | HER2             |                   |                |                      |                        |                   |                      |                                | HER3             |               |               |
|--|------------------|-------------------|----------------|----------------------|------------------------|-------------------|----------------------|--------------------------------|------------------|---------------|---------------|
|  | Breast<br>(n=25) | Bladder<br>(n=16) | Lung<br>(n=26) | Colorectal<br>(n=12) | Biliary tract<br>(n=9) | Cervical<br>(n=5) | Endometrial<br>(n=7) | Gastro-<br>esophageal<br>(n=5) | Ovarian<br>(n=4) | NOS<br>(n=16) | NOS<br>(n=16) |
| Patients continuing on treatment,<br>n (%) | 1 (4)            | 1 (6.2)           | 1 (3.8)        | 0 (0)                | 1 (11.1)               | 2 (40)            | 1 (14.3)             | 0 (0)                          | 1 (25)           | 2 (12.5)      | 0 (0)         |
| Treatment discontinuation, n (%)           | 24 (96.0)        | 15 (93.8)         | 25 (96.2)      | 12 (100)             | 8 (88.9)               | 3 (60.0)          | 6 (85.7)             | 5 (100)                        | 3 (75.0)         | 14 (87.5)     | 16 (100)      |
| Death                                      | 0 (0)            | 1 (6.3)           | 0 (0)          | 0 (0)                | 1 (11.1)               | 0 (0)             | 0 (0)                | 0 (0)                          | 0 (0)            | 0 (0)         | 1 (6.3)       |
| Disease progression                        | 22 (88.0)        | 12 (75.0)         | 18 (69.2)      | 11 (91.7)            | 5 (55.6)               | 2 (40.0)          | 4 (57.1)             | 4 (80.0)                       | 2 (50.0)         | 9 (56.3)      | 15 (93.8)     |
| Clinical deterioration                     | 0 (0)            | 0 (0)             | 3 (11.5)       | 0 (0)                | 1 (11.1)               | 0 (0)             | 0 (0)                | 1 (20.0)                       | 1 (25.0)         | 2 (12.5)      | 0 (0)         |
| Adverse Event                              | 0 (0)            | 0 (0)             | 2 (7.7)        | 0 (0)                | 1 (11.1)               | 0 (0)             | 1 (14.3)             | 0 (0)                          | 0 (0)            | 1 (6.3)       | 0 (0)         |
| Investigator Request                       | 0 (0)            | 1 (6.3)           | 0 (0)          | 1 (8.3)              | 0 (0)                  | 0 (0)             | 1 (14.3)             | 0 (0)                          | 0 (0)            | 2 (12.5)      | 0 (0)         |
| Withdrawal of consent                      | 2 (8.0)          | 0 (0)             | 2 (7.7)        | 0 (0)                | 0 (0)                  | 1 (20.0)          | 0 (0)                | 0 (0)                          | 0 (0)            | 0 (0)         | 0 (0)         |
| Lost to follow-up                          | 0 (0)            | 1 (6.3)           | 0 (0)          | 0 (0)                | 0 (0)                  | 0 (0)             | 0 (0)                | 0 (0)                          | 0 (0)            | 0 (0)         | 0 (0)         |
| Subjects ended study, n (%)                | 15 (60.0)        | 14 (87.5)         | 16 (61.5)      | 9 (75.0)             | 6 (66.7)               | 1 (20.0)          | 6 (85.7)             | 5 (100)                        | 3 (75.0)         | 7 (43.8)      | 14 (87.5)     |
| Death                                      | 12 (48.0)        | 13 (81.3)         | 13 (50.0)      | 8 (66.7)             | 6 (66.7)               | 1 (20.0)          | 5 (71.4)             | 3 (60.0)                       | 3 (75.0)         | 7 (43.8)      | 11 (68.8)     |
| Withdrawal of consent                      | 2 (8.0)          | 0 (0)             | 2 (7.7)        | 0 (0)                | 0 (0)                  | 0 (0)             | 1 (14.3)             | 0 (0)                          | 0 (0)            | 0 (0)         | 2 (12.5)      |
| Lost to follow-up                          | 1 (4.0)          | 1 (6.3)           | 1 (3.8)        | 1 (8.3)              | 0 (0)                  | 0 (0)             | 0 (0)                | 1 (20.0)                       | 0 (0)            | 0 (0)         | 1 (6.3)       |
| Other                                      | 0 (0)            | 0 (0)             | 0 (0)          | 0 (0)                | 0 (0)                  | 0 (0)             | 0 (0)                | 1 (20.0)                       | 0 (0)            | 0 (0)         | 0 (0)         |

NOS, not otherwise specified.

# Dynamic basis for dG·dT misincorporation via tautomerization and ionization

Isaac J. Kimsey<sup>1†\*</sup>, Eric S. Szymanski<sup>1\*</sup>, Walter J. Zahurancik<sup>2,3</sup>, Anisha Shakya<sup>4†</sup>, Yi Xue<sup>1†</sup>, Chia-Chieh Chu<sup>1</sup>, Bharathwaj Sathyamoorthy<sup>1†</sup>, Zucui Suo<sup>2,3</sup> & Hashim M. Al-Hashimi<sup>1,5</sup>

**Tautomeric and anionic Watson–Crick-like mismatches have important roles in replication and translation errors through mechanisms that are not fully understood. Here, using NMR relaxation dispersion, we resolve a sequence-dependent kinetic network connecting G·T/U wobbles with three distinct Watson–Crick mismatches: two rapidly exchanging tautomeric species ( $G^{\text{enol}}\cdot T/U \rightleftharpoons G\cdot T^{\text{enol}}/U^{\text{enol}}$ ; population less than 0.4%) and one anionic species ( $G\cdot T^-/U^-$ ; population around 0.001% at neutral pH). The sequence-dependent tautomerization or ionization step was inserted into a minimal kinetic mechanism for correct incorporation during replication after the initial binding of the nucleotide, leading to accurate predictions of the probability of dG·dT misincorporation across different polymerases and pH conditions and for a chemically modified nucleotide, and providing mechanisms for sequence-dependent misincorporation. Our results indicate that the energetic penalty for tautomerization and/or ionization accounts for an approximately  $10^{-2}$  to  $10^{-3}$ -fold discrimination against misincorporation, which proceeds primarily via tautomeric dG<sup>enol</sup>·dT and dG·dT<sup>enol</sup>, with contributions from anionic dG·dT<sup>-</sup> dominant at pH 8.4 and above or for some mutagenic nucleotides.**

In their paper describing the structure of the DNA double helix<sup>1</sup>, Watson and Crick proposed that if nucleotide bases adopted their energetically unfavourable tautomeric forms, mismatches (Fig. 1a) could pair up in a Watson–Crick (WC)-like geometry (Fig. 1b) and potentially give rise to spontaneous mutations. Decades later, it is well established that the replicative and translational machineries have a tight control over the WC geometry to discriminate against mismatches<sup>2–5</sup>. There is also evidence that both tautomeric<sup>6–13</sup> (Fig. 1b) and anionic<sup>7–9,14,15</sup> (Fig. 1c) WC-like mismatches can evade such fidelity checkpoints and give rise to errors in replication<sup>6,7</sup> and translation<sup>16</sup>. Although they are central to the fidelity of information transfer in molecular biology, and despite growing evidence that spontaneous mutations are involved in cancer-causing alterations<sup>17</sup>, the existence of WC-like mismatches and their contribution to replication and translation errors have not yet been definitively established.

Tautomeric and anionic mismatches exist in a variety of chemical forms (Extended Data Fig. 1). For example, WC-like G·T/U mismatches can form when either the guanine ( $G^{\text{enol}}\cdot T/U$  and  $G\cdot T/U$ ) or the thymidine or uridine ( $G\cdot T^{\text{enol}}/U^{\text{enol}}$  and  $G\cdot T^-/U^-$ ) base assumes a rare enolic (Fig. 1b) or anionic (Fig. 1c) form. Although it remains unclear which WC-like mismatch contributes to replication and translation errors, factors that stabilize different forms (for example, changes in pH<sup>7,8,14,18</sup> and chemical modifications<sup>19</sup>) have been shown to increase the probability of misincorporation<sup>14,20</sup>. Misincorporation probabilities can also vary markedly with sequence context, through mechanisms that are still poorly understood<sup>21,22</sup>. The resolution of these different WC-like mismatches and their chemical dynamics is key to the elucidation of their potential roles in replication, transcription and translation errors. However, this presents a formidable challenge for current biophysical methods because these mismatches differ only in the placement of a single proton and a  $\pi$ -bond (Fig. 1b, c and Extended Data Fig. 1). Protons are generally invisible to X-ray

crystallography and cryo-EM<sup>12</sup>, and consequently it has not been possible to unambiguously resolve the identity of WC-like mismatches captured in the active sites of polymerases<sup>6,7,15,23</sup> and the ribosome decoding site<sup>9,24</sup>. Moreover, WC-like mismatches are predicted to exist in rapid tautomeric equilibria ( $G^{\text{enol}}\cdot T/U \rightleftharpoons G\cdot T^{\text{enol}}/U^{\text{enol}}$ )<sup>25,26</sup> (Fig. 1b, c and Extended Data Fig. 1), making them exceptionally difficult to capture experimentally.

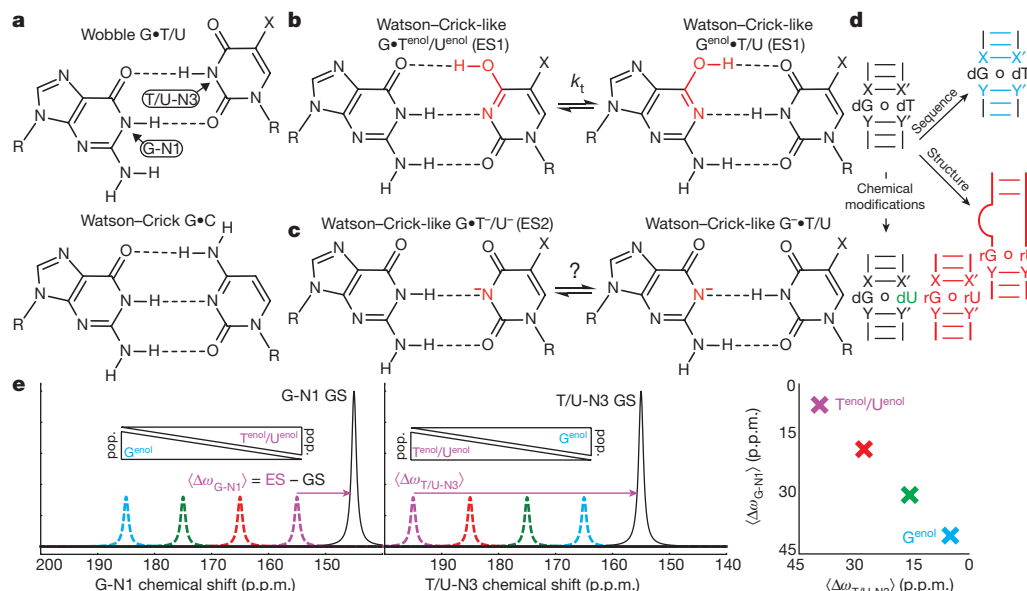
Techniques based on NMR relaxation dispersion<sup>27–29</sup> enable the characterization of low-abundance, short-lived conformational states, known as ‘excited states’ (ESs), in biomolecules<sup>30</sup>. Recently we used these techniques to provide evidence that wobble G·T/U mismatches exist in dynamic equilibrium with tautomeric (ES1) and anionic (ES2) WC-like mismatches in DNA and RNA duplexes<sup>8,31</sup>. The chemical shifts measured for guanine N1 (G-N1) and thymidine/uridine N3 (T/U-N3) in tautomeric ES1 were consistent with  $G^{\text{enol}}\cdot T/U$ , but were partially skewed towards  $G\cdot T^{\text{enol}}/U^{\text{enol}}$ . This was interpreted as evidence for a rapid (on the chemical-shift timescale) equilibrium between a major  $G^{\text{enol}}\cdot T/U$  and a minor  $G\cdot T^{\text{enol}}/U^{\text{enol}}$  species<sup>8</sup>. The anionic ES2 was detectable only at high pH (7.8 or above), and was heavily skewed in favour of  $G\cdot T^-/U^-$  with no evidence of  $G^- \cdot T/U$ . The roles of these various WC-like mismatches in replication and translation errors remain unknown. Here, by combining NMR relaxation dispersion and measurements of misincorporation rates, we resolved a kinetic network connecting two distinct tautomeric and one anionic WC-like mismatch, and established their relative contributions to dG·dTTP misincorporation.

## Tilting the tautomeric equilibrium

If ES1 does represent two tautomeric species in rapid equilibrium (Fig. 1b), it should be feasible to tilt the equilibrium ( $K_t = p_{G^{\text{enol}}\cdot T/U} / p_{G\cdot T^{\text{enol}}/U^{\text{enol}}}$ ) by changing the local sequence or structural context around the mismatch, or by modification of the bases (Fig. 1d). This in turn should lead to

<sup>1</sup>Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>2</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA. <sup>3</sup>The Ohio State Biochemistry Program, The Ohio State University, Columbus, Ohio 43210, USA. <sup>4</sup>Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>5</sup>Department of Chemistry, Duke University, Durham, North Carolina 27710, USA. <sup>†</sup>Present addresses: Nymirum, Durham, North Carolina 27713, USA (I.J.K.); Institute of Basic Science, Center for Soft and Living Matter, Ulsan, South Korea (A.S.); School of Life Sciences, Tsinghua University, Beijing, China (Y.X.); Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhopal 462 066, India (B.S.).

\*These authors contributed equally to this work.



**Figure 1 | Tilting the rapid tautomeric equilibria in excited-state WC-like mismatches.** **a–c**, Chemical structures of ground state wobble (**a**) and excited-state tautomeric (**b**) and anionic (**c**) WC-like G•T/U mismatches. X = H and CH<sub>3</sub> in uridine and thymidine, respectively. **d**, Tilting the rapid tautomeric equilibria using sequence (cyan), structure (red) and

chemical (green) modifications. X–X' and Y–Y' denote WC base pairs adjacent to the G•T/U mismatches. **e**, Perturbations that differentially tilt the tautomeric equilibrium induce changes in the chemical shifts of G-N1 (left) and T/U-N3 (middle); a plot of  $\langle \Delta\omega_{G-N1} \rangle$  against  $\langle \Delta\omega_{T/U-N3} \rangle$  (right) is linear with a negative slope.

very specific changes in the chemical shifts ( $\omega$ ) of ES1 G-N1 and T/U-N3, which are provided as population-weighted average values denoted as  $\langle \omega \rangle$  over the two species (Fig. 1e, left). Tilting the equilibrium in favour of G<sup>enol</sup>•T/U should induce a downfield shift of  $\omega$  for ES1 G-N1, due to an increase in the population of deprotonated G<sup>enol</sup>, and an upfield shift of  $\omega$  for ES1 T/U-N3, due to a decrease in the population of deprotonated T<sup>enol</sup>/U<sup>enol</sup>, and vice versa (Fig. 1e, left). A plot of  $\langle \Delta\omega_{G-N1} \rangle$  against  $\langle \Delta\omega_{T/U-N3} \rangle$  is predicted to be linear (Fig. 1e, right), with a negative slope and intercept determined by the fundamental chemical shifts of the tautomeric species (see equation (1) in Supplementary Methods).

We measured <sup>15</sup>N relaxation dispersion for 5 dG•dT mismatches within distinct sequence contexts and for 13 rG•rU mismatches in 9 structurally unique non-coding RNAs (Fig. 2a and Extended Data Fig. 2a). Experiments were carried out at near-neutral pH (6.4–6.9) in order to reduce levels of the anionic ES2 below detection limits<sup>8</sup> (Extended Data Fig. 2b). The relaxation dispersion experiments measure spin-relaxation rates in the rotating frame ( $R_{1\rho}$ ) during a relaxation period in which a radiofrequency field is applied with variable offset ( $\Omega$  2 $\pi$ <sup>–1</sup>, in Hz) and power ( $\omega$  2 $\pi$ <sup>–1</sup>, in Hz), in order to suppress the chemical exchange contribution ( $R_{ex}$ ) to the transverse spin relaxation rate ( $R_2$ ) arising from chemical exchange between the energetically more stable ground state (GS) and the ES<sup>27,28</sup>.

We observed G-N1 and T/U-N3 relaxation dispersion consistent with WC-like ES1 exchange for all five dG•dT and eight rG•rU mismatches located within helical environments (Fig. 2b and Extended Data Fig. 3a), thus establishing their widespread occurrence in DNA and RNA. No relaxation dispersion was observed (Extended Data Fig. 3b) for rG•rU mismatches adjacent to apical loops, three-way junctions, or bulges (Extended Data Fig. 2a). This could be due to the lower abundance of WC-like mismatches when outside of the helical environment, although we cannot rule out the possibility that the exchange is orders of magnitude faster and beyond detection.

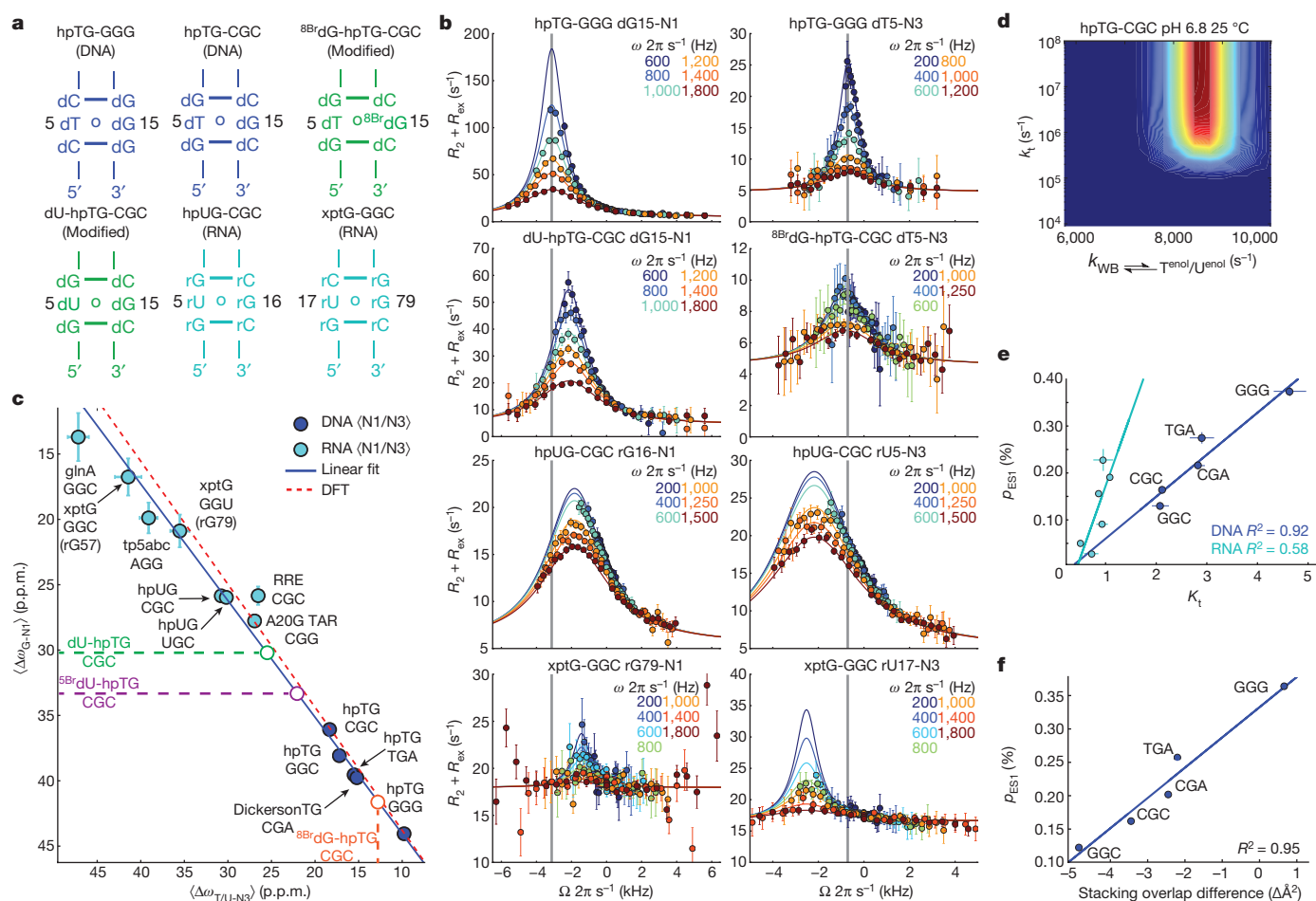
As predicted on the basis of variably tilting the G<sup>enol</sup>•T/U  $\rightleftharpoons$  G•T<sup>enol</sup>/U<sup>enol</sup> equilibrium (Fig. 1e, right), plotting the fitted  $\langle \Delta\omega_{G-N1} \rangle$  and  $\langle \Delta\omega_{T/U-N3} \rangle$  values obtained from two-state analysis (GS  $\rightleftharpoons$  ES1) of the relaxation dispersion profiles (Fig. 2b, Extended Data Fig. 3a and Supplementary Table 1) formed a line with a negative slope (Fig. 2c). As a negative control, the corresponding GS G-N1 and T/U-N3

chemical shifts were plotted and no correlation was observed (Extended Data Fig. 3c). We confirmed these linear trends using chemical modifications that tilt the tautomeric equilibrium towards enolic dT (dG•dU and dG•<sup>5</sup>BrdU) or enolic dG (<sup>8</sup>BrdG•dT) (Fig. 2c, Extended Data Fig. 3a and Supplementary Discussion 1).

### Sequence-dependent G<sup>enol</sup>•T/U $\rightleftharpoons$ G•T<sup>enol</sup>/U<sup>enol</sup>

A linear fit of the values obtained by plotting  $\langle \Delta\omega_{G-N1} \rangle$  against  $\langle \Delta\omega_{T/U-N3} \rangle$ , assuming physically reasonable ranges, yielded fundamental chemical shifts for the tautomeric species that are in excellent agreement with values predicted by DFT calculations (Fig. 2c and Supplementary Tables 2, 3)<sup>8</sup>. The tautomeric equilibria (Supplementary Table 2) obtained from this analysis and from refitting the relaxation dispersion data using a three-state model with linear topology<sup>25</sup> (wobble  $\rightleftharpoons$  G•T<sup>enol</sup>/U<sup>enol</sup>  $\rightleftharpoons$  G<sup>enol</sup>•T/U) are slightly tilted in favour of dG<sup>enol</sup>•dT in DNA ( $K_t = 2.1$ –4.6), whereas the populations of rG<sup>enol</sup>•rU and rG•rU<sup>enol</sup> are more comparable in RNA ( $K_t = 0.5$ –1.1). These differences may be attributed to the electron-donating methyl group in dT, which destabilizes dT<sup>enol</sup> relative to rU<sup>enol</sup> (ref. 32). The relaxation dispersion data also enabled us to estimate a lower bound for the fast tautomeric exchange rate of  $k_t = k_{G^{enol} \rightarrow T^{enol}} + k_{T^{enol} \rightarrow G^{enol}} > \approx 500,000$ –1,000,000 s<sup>–1</sup>, faster than previously measured exchange processes by similar RD methods<sup>33</sup> (Fig. 2d and Extended Data Fig. 4), and a transition-state barrier for the conversion of G•T<sup>enol</sup>/U<sup>enol</sup>  $\rightarrow$  G<sup>enol</sup>•T/U of <9–10 kcal mol<sup>–1</sup> (using  $k_B T h^{-1}$  as the pre-exponential factor<sup>34</sup> and  $\kappa = 1$  as the transmission coefficient), which is in good agreement with values reported using computational methods (around 11.5 kcal mol<sup>–1</sup>)<sup>25</sup>. These results establish the existence of G•T<sup>enol</sup>/U<sup>enol</sup> and G<sup>enol</sup>•T/U in an ultra-fast equilibrium, each of which can potentially contribute to replication and translation errors.

Notably, the exchange parameters vary markedly with sequence context (Supplementary Table 1). The ES1 population ( $p_{ES1} = p_{G^{enol}} + p_{T^{enol}/U^{enol}}$ ) varies 3-fold in DNA and 8-fold in RNA, whereas the forward ( $k_{GS \rightarrow ES1}$ ) and reverse ( $k_{ES1 \rightarrow GS}$ ) rate constants vary by 4- and 5-fold, respectively, for DNA, and by 38- and 6-fold, respectively, for RNA (Supplementary Table 1). A linear correlation is observed between  $p_{ES1}$  and  $K_t$  (Fig. 2e and Supplementary Table 2), indicating that the G<sup>enol</sup>•T/U population dominates these variations with sequence and structural context. In DNA, these variations can potentially be explained by the sequence-specific



**Figure 2 | Resolving rapidly interconverting tautomers.** **a**, Representative hairpin (hp) DNA, RNA and chemically modified constructs. The name denotes the mismatch and sequence context (5'-3'). **b**, Representative G-N1 and T/U-N3 relaxation dispersion profiles (pH 6.8–6.9 and 25 °C). Best fits to the Bloch–McConnell (B–M) equations are shown. **c**, A plot of ES1 chemical shift differences,  $\langle \Delta\omega_{G-N1} \rangle$  against  $\langle \Delta\omega_{T/U-N3} \rangle$ , in ES1 as measured by NMR relaxation dispersion. The blue line indicates the fit to the  $\Delta\omega$  values using the fundamental tautomer chemical shifts as variables (Supplementary Methods). The red line indicates predictions from density

changes in stacking with immediate neighbours that accompany the transition from the wobble to the Watson–Crick geometry (Fig. 2f). For example, GGG has the highest  $p_{\text{ES1}}$  and is predicted to gain stacking overlap, whereas CGG has the lowest  $p_{\text{ES1}}$  and is predicted to lose stacking overlap. Similar sequence-dependent effects have been reported for lesion repair by methyltransferases<sup>35</sup>. Notably, dG dominates the changes in stacking, potentially explaining the stronger sequence dependence of the  $G^{\text{enol}} \cdot T/U$  population compared with that of  $G \cdot T^{\text{enol}}/U^{\text{enol}}$ .

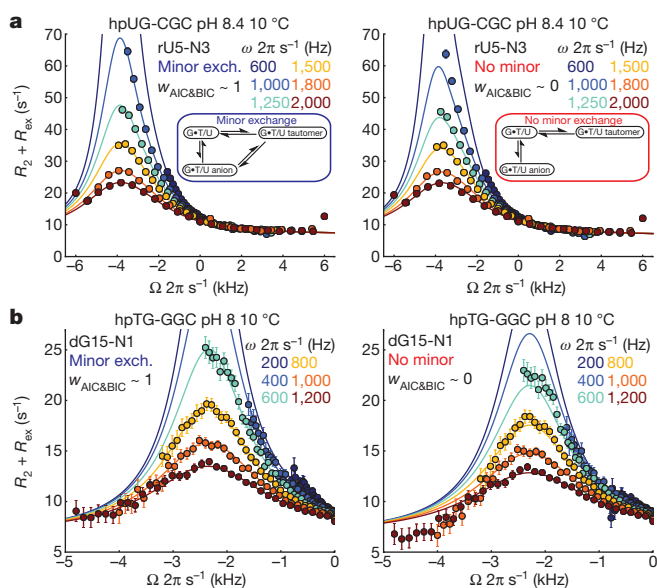
### Sequence-dependent anion equilibria

Next, we examined whether anionic WC-like  $G \cdot T^{\text{enol}}/U^{\text{enol}}$  (Fig. 1c) also form commonly in DNA and RNA, and whether anionic  $G^{\text{enol}} \cdot T/U$  remains undetectable under these different environments. We measured relaxation dispersion at high pH (7.8 and above) for  $G \cdot T/U$  mismatches in a subset of our RNA (Fig. 3a and Extended Data Fig. 5a) and DNA (Fig. 3b and Extended Data Fig. 5b) constructs. In all cases, we observed relaxation dispersion consistent with wobble  $\rightleftharpoons$  anion exchange, thus establishing the robustness of this process across different sequence contexts (Fig. 3a, b and Extended Data Fig. 5). Three-state fitting of the relaxation dispersion data, assuming a starlike topology (Extended Data Fig. 6a), yielded large  $\Delta\omega_{T/U-N3(\text{ES2})}$  values of approximately 55 p.p.m. and much smaller  $\Delta\omega_{G-N1(\text{ES2})}$  values of approximately 5 p.p.m., consistent with a dominant  $G \cdot T^{\text{enol}}/U^{\text{enol}}$  species

functional theory. **d**, Lower bounds for the rate of tautomeric  $G^{\text{enol}} \cdot T/U \rightleftharpoons G \cdot T^{\text{enol}}/U^{\text{enol}}$  exchange. Contour plots showing scaled  $\bar{\chi}^2$  weights for combinations of  $k_1$  against  $k_{\text{Tenol/Uenol}}$ ; red indicates the better fit. **e**, A plot of tautomeric ES1 population against  $K_t$  for DNA (blue,  $n = 5$ ) and RNA (cyan,  $n = 6$ ) constructs determined at pH 6.9 and 25 °C. **f**, Plot of ES1 population against change in stacking overlap ( $\Delta\bar{A}^2 = \bar{A}^2(\text{WC}) - \bar{A}^2(\text{WB})$ ) between wobble and Watson–Crick-like mismatches (pH 6.9, 25 °C) for five DNA sequence contexts (Supplementary Methods). Error bars in **b**, **c** and **e** reflect experimental uncertainty (one s.d., Supplementary Methods).

and with no evidence of  $G^{\text{enol}} \cdot T/U$ . Again, we observe strong sequence-specific variations in the ES2 population ( $p_{\text{ES2}}$ ) and in the values of  $k_{\text{GS} \rightarrow \text{ES2}}$  and  $k_{\text{ES2} \rightarrow \text{GS}}$  across different temperatures and pH values (Supplementary Tables 4, 5).

A previous study<sup>8</sup> showed that the emergence of anionic ES2 at high pH values was accompanied by unexpected changes in the tautomeric chemical shifts of ES1. Similar deviations are observed here for both RNA and DNA (Supplementary Table 4). We postulated that ‘minor’ exchange<sup>36</sup> between ES1 and ES2 could ‘mix’ their chemical shifts and give rise to such deviations (Extended Data Fig. 6a, b and Supplementary Table 4). Indeed, all five relaxation dispersion profiles with unusual ES1 chemical shifts showed a statistically significant improvement when the data was fitted to a three-state model with minor exchange in a triangular rather than a starlike topology (Fig. 3a, b, Extended Data Fig. 6c, d and Supplementary Tables 4–6). The resulting ES1  $^{15}\text{N}$  rG-N1 and rU-N3 chemical shifts vary less significantly with pH (Extended Data Fig. 6e and Supplementary Table 4) and the rate constants ( $k_{\text{ES1} \rightarrow \text{ES2}}$  and  $k_{\text{ES2} \rightarrow \text{ES1}}$ ) exhibit the expected temperature dependence (Extended Data Fig. 6f), neither of which would be expected if the data were being spuriously overfitted. It should be noted that limited or poor quality relaxation dispersion data can make it difficult to resolve different topologies<sup>37</sup> (Supplementary Table 6).



**Figure 3 | Three-state exchange with triangular topology and minor exchange between tautomeric and anionic WC-like excited states.** **a, b,** Comparison of three-state B–M fit with triangular (left) and starlike (right) topologies with the relaxation dispersion profiles measured in hpUG-CGC RNA (**a**) and hpTG-GGC DNA (**b**). Statistical Akaike's information criterion and Bayesian information criterion weights ( $w_{\text{AIC}}$  and  $w_{\text{BIC}}$ , respectively) comparing starlike and triangular topologies are shown. Error bars reflect experimental uncertainty (one s.d., Supplementary Methods).

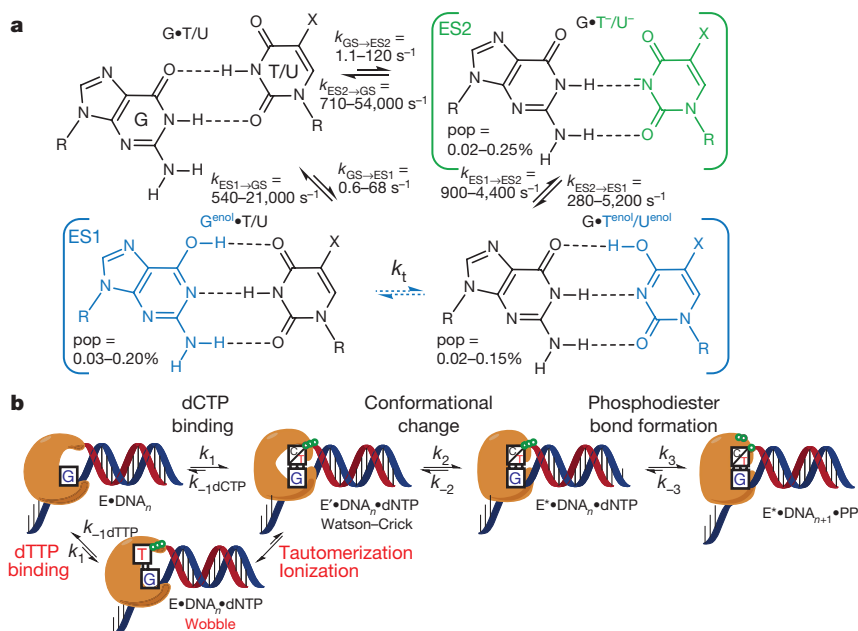
### Tautomerization and ionization during misincorporation

dG•dT misincorporation is the most frequent base-substitution error committed by high-fidelity DNA polymerases, with a misincorporation frequency  $F_{\text{pol}}$  of  $10^{-4}$ – $10^{-5}$  for most studied polymerases<sup>38,39</sup> (where

$F_{\text{pol}} = (k_{\text{pol}}/K_{\text{d}})_{\text{incorrect}} / (k_{\text{pol}}/K_{\text{d}})_{\text{correct}}$ ;  $k_{\text{pol}}$  is the maximum rate of nucleotide incorporation and  $K_{\text{d}}$  is the apparent nucleotide equilibrium dissociation constant)—that is, an error is committed at a frequency of 1 in every  $10^4$ – $10^5$  nucleotide incorporations. Differences in apparent nucleotide binding affinities ( $K_{\text{d}}^{-1}(\text{incorrect})/K_{\text{d}}^{-1}(\text{correct})$ ) account for a factor of only around  $10^{-1}$ – $10^{-2}$  in discrimination<sup>40</sup>, whereas differences in the polymerization rates ( $k_{\text{pol}}^{\text{incorrect}}/k_{\text{pol}}^{\text{correct}}$ ) account for a factor of approximately  $10^{-3}$ .

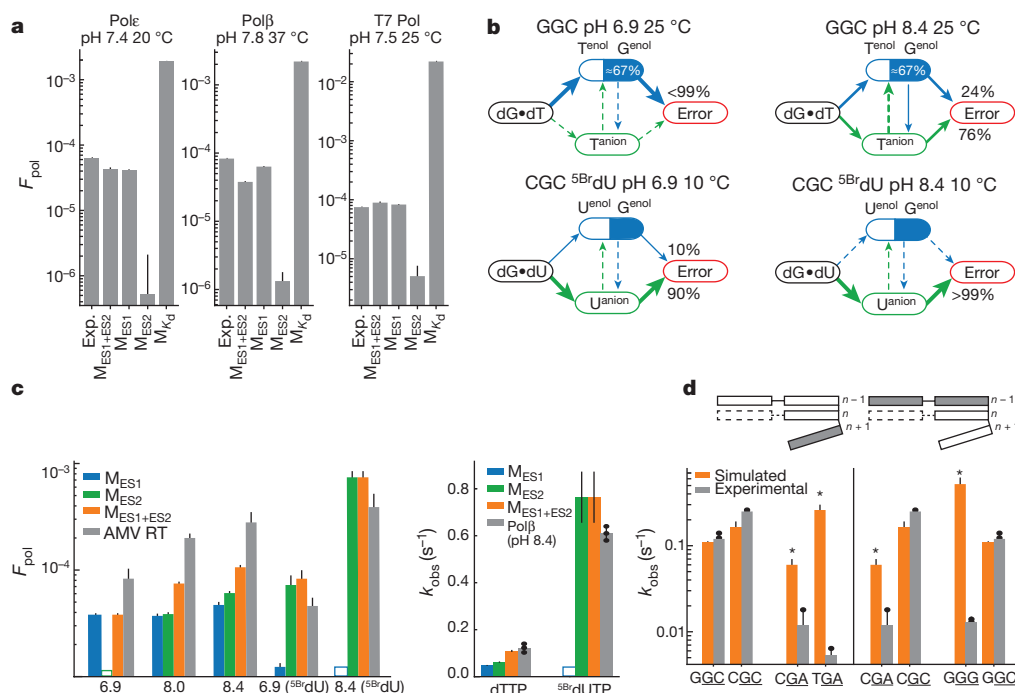
The mechanisms that lower the values of  $k_{\text{pol}}^{\text{incorrect}}$  relative to  $k_{\text{pol}}^{\text{correct}}$  are still poorly understood. Several decades ago, Topal and Fresco postulated that the frequency of tautomerization may be an important determinant of misincorporation probability<sup>11</sup>. Notably, the population of the tautomeric species (around  $10^{-3}$ ) is comparable to the values of  $k_{\text{pol}}^{\text{incorrect}}/k_{\text{pol}}^{\text{correct}}$ . In addition, the rate at which the wobble dG•dT forms either the WC-like tautomeric ( $k_{\text{GS} \rightarrow \text{ES}1} = 0.3$ – $10 \text{ s}^{-1}$ ; Supplementary Tables 1, 5) or anionic ( $k_{\text{GS} \rightarrow \text{ES}2} = 1.1$ – $124 \text{ s}^{-1}$ ; Supplementary Tables 1, 5) mismatches (Fig. 4a) is comparable to the values of  $k_{\text{pol}}^{\text{incorrect}}$  ( $0.16$ – $1.16 \text{ s}^{-1}$ ) measured for incorrect dG•dTTP or dGTP•dT misincorporation<sup>22,39</sup>, whereas it is up to approximately 1000-fold slower than  $k_{\text{pol}}^{\text{correct}}$  ( $25$ – $275 \text{ s}^{-1}$ ) measured for correct dG•dCTP or dGTP•dCT misincorporation<sup>22,39</sup>. If the formation of WC-like dG•dT mismatches (Fig. 4a) is required for misincorporation after the initial binding of dNTP in a wobble conformation, it could provide a mechanism for lowering  $k_{\text{pol}}^{\text{incorrect}}$  relative to  $k_{\text{pol}}^{\text{correct}}$ . Indeed, previous studies have shown that DNA polymerases cannot undergo the necessary conformational changes needed for catalysis when dG•dT is in a wobble conformation<sup>7</sup> and all available structures of catalytically active polymerases with bound mismatches within the active site feature WC-like dG•dT or dA•dC geometries<sup>6,7</sup>. Similarly, WC-like rG•rU mismatches have been shown to form in the first and second codon positions of catalytically active ribosomes<sup>9</sup>, in which wobbles are typically rejected<sup>5</sup>, which may help to explain translational error hotspots<sup>41</sup>.

To examine this possibility, we built a kinetic model for dG•dTTP misincorporation by inserting a tautomerization or ionization step



**Figure 4 | Kinetic mechanism of dG•dT misincorporation.** **a,** Exchange between wobble G•T/U mispair (top left), rapidly interconverting WC-like tautomers ( $\text{G}^{\text{enol}}\cdot\text{T}/\text{U} \rightleftharpoons \text{G}^{\text{anion}}\cdot\text{T}/\text{U}$ , bottom), and anionic WC-like G•T-/U- (top right). Exchange between anionic G•T-/U- and a low-abundance, short-lived anionic G•T/U, or other non-WC species that cannot be detected by relaxation dispersion, cannot be ruled out. WC-like G•T/U populations and ranges recorded at pH 6.4–8.9 and 10–25 °C

(Supplementary Tables 1, 5). **b,** Minimal kinetic mechanism for polymerization<sup>42</sup>. Incorporation of an incorrect dTTP includes an additional tautomerization or ionization step, allowing for the formation of a Watson–Crick-like dG•dT mismatch. Discriminatory steps are in red. E, open polymerase conformation; E', closed conformation; E\*, closed, catalytically competent conformation.



**Figure 5 | Measured versus predicted misincorporation probabilities and rates.** **a**,  $F_{\text{pol}}$  as measured experimentally for dTTP•dG misincorporation for human DNA polymerase  $\epsilon$ , rat DNA polymerase  $\beta$  and T7 DNA polymerase, with values simulated using  $M_{\text{ES1}}$ ,  $M_{\text{ES2}}$ ,  $M_{\text{ES1}} + \text{ES2}$  and  $M_{\text{Kd}}$  (error bars represent one s.d., Supplementary Methods). **b**, Flux pathways for dT•dG(GGC) (top) and  $5\text{BrdU} \cdot \text{dG}(\text{CGC})$  (bottom), at different temperatures and pH values. **c**, Left, measured and simulated  $F_{\text{pol}}$  values for dTTP/ $5\text{BrdU}$  in AMV RT<sup>14</sup>. Right, measured and simulated  $k_{\text{obs}}$  values for dTTP/ $5\text{BrdU}$  misincorporation for human

DNA polymerase  $\beta$ . Error is s.d. of  $n = 3$  biological replicates for kinetic assays, or previously published error for AMV RT<sup>14</sup>. Error for kinetic simulations is described in the Supplementary Methods. **d**, Measured and simulated  $k_{\text{obs}}$  for dTTP misincorporation for human DNA polymerase  $\beta$  in different sequence contexts. The asterisk indicates that ES2 exchange rates were extrapolated (Supplementary Methods). Error is s.d. of  $n = 3$  biological replicates for kinetic assays; error for kinetic simulations is described in the Supplementary Methods.

(Fig. 4a) after initial nucleotide binding in a wobble conformation and before the pre-chemistry conformational change in the existing minimal kinetic model for correct incorporation<sup>42</sup> (Fig. 4b). All other steps, including the pre-chemistry conformational change and the phosphodiester bond formation, are assumed to have identical kinetic parameters as measured for correct nucleotide incorporation<sup>39,43–46</sup> (Supplementary Table 7). The model assumes that misincorporation directly from the wobble conformation is negligible and that the tautomerization and ionization rates measured in duplex DNA by NMR approximate the rates in the polymerase active site. We tested models (Extended Data Fig. 7) in which the tautomeric ( $M_{\text{ES1}}$ ), the anionic ( $M_{\text{ES2}}$ ), or both ( $M_{\text{ES1}} + \text{ES2}$ ) species can be misincorporated, as well as models that excluded the triangular network altogether ( $M_{\text{Kd}}$ ).

Notably, the most general  $M_{\text{ES1}} + \text{ES2}$  model robustly predicts the measured  $F_{\text{pol}}$  values for three polymerases (T7, polymerase  $\epsilon$  and polymerase  $\beta$ ) that have varying rate limiting steps and microscopic rate constants (Fig. 5a, Extended Data Fig. 8 and Supplementary Table 7). Similar results are obtained with  $M_{\text{ES1}}$  under these neutral conditions, in which the ES2 population is negligible ( $<10^{-6}$  at pH 6.9) (Fig. 5a and Extended Data Fig. 8). By contrast,  $M_{\text{ES2}}$  consistently underestimates  $F_{\text{pol}}$  by one to two orders of magnitude<sup>47</sup>, whereas  $M_{\text{Kd}}$  overestimates  $F_{\text{pol}}$  by one to two orders of magnitude<sup>40</sup> (Fig. 5a). Variants of the  $M_{\text{Kd}}$  model, in which only preformed tautomeric dNTP with populations of  $10^{-4}$ – $10^{-5}$  bind in a productive WC-like geometry, overestimate  $k_{\text{pol}}$  and  $K_d$  by several orders of magnitude (data not shown). These data indicate that the formation of tautomeric WC-like  $\text{dG}^{\text{enol}} \cdot \text{dT}$  and  $\text{dG} \cdot \text{dT}^{\text{enol}}$  at a population of around 0.1% can account for the approximately  $10^2$ – $10^3$ -fold lower value of  $k_{\text{pol}}^{\text{incorrect}}$  relative to  $k_{\text{pol}}^{\text{correct}}$ , and that at neutral pH more than 99% of misincorporation proceeds via the tautomeric species, which form predominantly via direct exchange from the wobble (Fig. 5b).

## Impact of pH, modifications and sequence

We also examined whether the  $M_{\text{ES1}} + \text{ES2}$  model can reproduce the dependence of the misincorporation probability on pH, base modification and sequence.  $M_{\text{ES1}} + \text{ES2}$  accurately predicts the approximately threefold increase in misincorporation probability observed with increasing pH (Fig. 5c, left). This can be attributed to an increase in the population of  $\text{dG} \cdot \text{dT}^-$ , which accounts for more than 70% of the net misincorporation at pH 8.4 (Fig. 5b, c). By contrast,  $M_{\text{ES1}}$  fails to predict this increase in misincorporation probability (Fig. 5c, left;  $M_{\text{Kd}}$  not shown owing to absence of pH-dependent  $K_d$  values). At high pH values, the tautomeric and anionic species have comparable populations, and there is significant flux (greater than 20%) towards both tautomeric and anionic species through the indirect minor exchange pathway (Fig. 5b). In this manner, the contributions of the tautomeric and anionic species to misincorporation are coupled.

$M_{\text{ES1}} + \text{ES2}$  and NMR relaxation dispersion measurements also accurately predict  $F_{\text{pol}}$  and  $k_{\text{obs}}$  for 5-bromo-2'-deoxyuridine triphosphate ( $5\text{BrdUTP}$ ) (Fig. 5c). This includes a sharper, approximately sixfold increase in  $F_{\text{pol}}^{\text{dG} \cdot 5\text{BrdUTP}}$  measured for avian myeloblastosis virus reverse transcriptase (AMV RT) when increasing the pH from 6.9 to 8.4 (Fig. 5c, left). This can be attributed to the lower  $\text{pK}_a$  of  $\text{dG} \cdot 5\text{BrdU}^-$  ( $\text{pK}_a \sim 9$ ) (ref. 14) relative to  $\text{dG} \cdot \text{dT}^-$  ( $\text{pK}_a \sim 11.8$ ) (ref. 8). We further verified the robustness of these predictions by measuring  $k_{\text{obs}}^{\text{dG} \cdot 5\text{BrdUTP}}$  and  $k_{\text{obs}}^{\text{dTTP}}$  for human DNA polymerase  $\beta$  at high pH (8.4). The model accurately predicts the approximately fourfold enhancement in  $k_{\text{obs}}^{\text{dG} \cdot 5\text{BrdUTP}}$  relative to  $k_{\text{obs}}^{\text{dTTP}}$  (Fig. 5c, right). Again,  $M_{\text{ES1}}$  fails to predict these variations (Fig. 5c). Indeed, at both neutral and high pH values,  $5\text{BrdUTP}$  is predicted to be predominantly misincorporated via the more populated  $\text{dG} \cdot \text{dT}^-$  (Fig. 5b, c). These data indicate that misincorporation due to  $\text{dG} \cdot \text{dT}^-$  can dominate at pH values of 8.4 or above, or for chemically modified nucleotides at neutral pH.

Importantly, owing to the sequence-dependence of tautomerization and ionization,  $M_{ES1+ES2}$  also predicts sequence-specific variations in values of  $F_{pol}$  of approximately eightfold at pH 8.4 (Fig. 5d). Comparable (fivefold) sequence-specific variations have been reported previously<sup>21</sup>. We tested these predictions using human DNA polymerase  $\beta$  at pH 8.4 for nine different sequence contexts (Supplementary Table 8). Whereas  $k_{obs}^{dG \cdot dCTP}$  varied weakly (less than 1.2-fold) with sequence,  $k_{obs}^{dG \cdot dTTP}$  varied approximately 45-fold (Extended Data Fig. 9); larger changes were observed when the base pair was altered at the  $n - 1$  position, which stacks with dG•dTTP in the polymerase active site (Fig. 5d). Although the  $M_{ES1+ES2}$  predictions slightly underestimate the sequence-specific variations in  $k_{obs}^{dG \cdot dTTP}$ , this is not too surprising considering that other reaction steps could also vary with sequence. The predictions do recapitulate the lower  $k_{obs}^{dG \cdot dTTP}$  for CGA and comparable values for GGC and CGC (Fig. 5d). Notably, the two major outliers (TGA and GGG) arise primarily because of a large ES2 population. It is likely that the polymerase environment, including the absence of base pairs at the  $n + 1$  position (Fig. 5d), can influence the sequence-specific dependence of tautomerization and ionization, and consequently influence misincorporation.

Our data indicate that the formation of WC-like anionic and tautomeric mismatches help to determine the frequency of dG•dT misincorporation and its dependence on pH, chemical modifications and possibly sequence. Our analysis indicates that  $F_{pol}$  is determined primarily by the ES1 population, and that considerable reductions in  $k_{ex} = k_{GS \rightarrow ES1} + k_{ES1 \rightarrow GS}$ , outside of the range detected here, would be required to substantially reduce  $F_{pol}$  (Extended Data Fig. 10). Although it is likely that differences in the active site environment of the polymerase will tune tautomerization and ionization dynamics, the robustness of the predictions across different polymerases, pH conditions, and modified nucleotides suggests that it will not cause substantial perturbations relative to the broad kinetic range examined here. Indeed, very small differences in tautomerization and ionization dynamics are observed for DNA and RNA, which have different helical structures and stabilities. It is possible that tautomerization and ionization are dominated by the energetics of hydrogen bonding and proton transfer, and that the natural grip over WC geometry in the double helix is similar to that achieved by the polymerase in the context of an isolated dNTP paired to the template. Other mechanisms may be applicable for purine–purine mismatches for which alterations in the active site have been proposed rather than the adoption of a WC-like base pair<sup>43,48</sup>. The approach presented here can be applied to examine the roles of other tautomeric and anionic mismatches in replication, transcription, translation and DNA repair<sup>49</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Data Availability** The data that support the findings of this study are available upon reasonable request from the corresponding authors.

**Received 10 June 2016; accepted 21 December 2017.**

**Published online 31 January 2018.**

- Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131 (1953).
- Oertell, K. *et al.* Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. *Proc. Natl Acad. Sci. USA* **113**, E2277–E2285 (2016).
- Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* **82**, 203–236 (2013).
- Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219 (2002).
- Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256–259 (2012).
- Wang, W., Hellinga, H. W. & Beese, L. S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA* **108**, 17644–17648 (2011).
- Bebenek, K., Pedersen, L. C. & Kunkel, T. A. Replication infidelity via a mismatch with Watson–Crick geometry. *Proc. Natl Acad. Sci. USA* **108**, 1862–1867 (2011).

- Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320 (2015).
- Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. & Yusupova, G. Structural insights into the translational infidelity mechanism. *Nat. Commun.* **6**, 7251 (2015).
- Topal, M. D. & Fresco, J. R. Base pairing and fidelity in codon–anticodon interaction. *Nature* **263**, 289–293 (1976).
- Topal, M. D. & Fresco, J. R. Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285–289 (1976).
- Loveland, A. B., Demo, G., Grigorieff, N. & Korostelev, A. A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113–117 (2017).
- Singh, V., Fedele, B. I. & Essigmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **21**, 1–13 (2015).
- Yu, H., Eritja, R., Bloom, L. B. & Goodman, M. F. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *J. Biol. Chem.* **268**, 15935–15943 (1993).
- Koag, M. C., Nam, K. & Lee, S. The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase  $\beta$ . *Nucleic Acids Res.* **42**, 11233–11245 (2014).
- Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. & Yusupova, G. New structural insights into translational miscoding. *Trends Biochem. Sci.* **41**, 798–814 (2016).
- Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330–1334 (2017).
- Eckert, K. A. & Kunkel, T. A. Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J. Biol. Chem.* **268**, 13462–13471 (1993).
- Warren, J. J., Forsberg, L. J. & Beese, L. S. The structural basis for the mutagenicity of O<sup>6</sup>-methyl-guanine lesions. *Proc. Natl Acad. Sci. USA* **103**, 19701–19706 (2006).
- Dosanjh, M. K., Galeros, G., Goodman, M. F. & Singer, B. Kinetics of extension of O<sup>6</sup>-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry* **30**, 11595–11599 (1991).
- Mendelman, L. V., Boosalis, M. S., Petruska, J. & Goodman, M. F. Nearest neighbor influences on DNA polymerase insertion fidelity. *J. Biol. Chem.* **264**, 14415–14423 (1989).
- Lee, H. R. & Johnson, K. A. Fidelity of the human mitochondrial DNA polymerase. *J. Biol. Chem.* **281**, 36236–36240 (2006).
- Xia, S. & Konigsberg, W. H. Mispairs with Watson–Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. *Protein Sci.* **23**, 508–513 (2014).
- Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721–732 (2002).
- Nomura, K. *et al.* DFT calculations on the effect of solvation on the tautomeric reactions for wobble gua–thy and canonical gua–cyt base-pairs. *J. Mod. Phys.* **4**, 422–431 (2013).
- Brovarets', O. O. & Hovorun, D. M. The nature of the transition mismatches with Watson–Crick architecture: the G•T or G•T\* DNA base mispair or both? A QM/QTAIM perspective for the biological problem. *J. Biomol. Struct. Dyn.* **33**, 925–945 (2015).
- Trott, O., Abergel, D. & Palmer, A. G., III. An average-magnetization analysis of  $R_{1\rho}$  relaxation outside of the fast exchange limit. *Mol. Phys.* **101**, 753–763 (2003).
- Korzhnev, D. M., Orekhov, V. Y. & Kay, L. E. Off-resonance  $R_{1\rho}$  NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J. Am. Chem. Soc.* **127**, 713–721 (2005).
- Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon  $R_{1\rho}$  NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 3818–3819 (2009).
- Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8**, 932–935 (2001).
- Szymanski, E. S., Kimsey, I. J. & Al-Hashimi, H. M. Direct NMR evidence that transient tautomeric and anionic states in dG•dT form Watson–Crick-like base pairs. *J. Am. Chem. Soc.* **139**, 4326–4329 (2017).
- Orozco, M., Hernández, B. & Luque, F. J. Tautomerism of 1-methyl derivatives of uracil, thymine, and 5-bromouracil. Is tautomerism the basis for the mutagenicity of 5-bromouridine? *J. Phys. Chem. B* **102**, 5228–5233 (1998).
- Ban, D. *et al.* Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J. Magn. Reson.* **221**, 1–4 (2012).
- Vallurupalli, P., Chakrabarti, N., Pomes, R. & Kay, L. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* **7**, 3602–3613 (2016).
- Delaney, J. C. & Essigmann, J. M. Effect of sequence context on O<sup>6</sup>-methylguanine repair and replication *in vivo*. *Biochemistry* **40**, 14968–14975 (2001).
- Trott, O. & Palmer, A. G., III. Theoretical study of  $R_{1\rho}$  rotating-frame and R2 free-precession relaxation in the presence of *n*-site chemical exchange. *J. Magn. Reson.* **170**, 104–112 (2004).
- Sekhar, A. *et al.* Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *eLife* **4**, e07296 (2015).

38. Kunkel, T. A. & Alexander, P. S. The base substitution fidelity of eukaryotic DNA polymerases. Mismatching frequencies, site preferences, insertion preferences, and base substitution by dislocation. *J. Biol. Chem.* **261**, 160–166 (1986).
39. Zahurancik, W. J., Klein, S. J. & Suo, Z. Significant contribution of the 3'→5' exonuclease activity to the high fidelity of nucleotide incorporation catalyzed by human DNA polymerase  $\epsilon$ . *Nucleic Acids Res.* **42**, 13853–13860 (2014).
40. Loeb, L. A. & Kunkel, T. A. Fidelity of DNA synthesis. *Annu. Rev. Biochem.* **51**, 429–457 (1982).
41. Zhang, J., Leong, K.-W., Johansson, M. & Ehrenberg, M. Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome. *Proc. Natl Acad. Sci. USA* **112**, 9602–9607 (2015).
42. Rothwell, P. J. & Waksman, G. Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71**, 401–440 (2005).
43. Tsai, Y. C. & Johnson, K. A. A new paradigm for DNA polymerase specificity. *Biochemistry* **45**, 9675–9687 (2006).
44. Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511–525 (1991).
45. Zahurancik, W. J., Klein, S. J. & Suo, Z. Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase  $\epsilon$ . *Biochemistry* **52**, 7041–7049 (2013).
46. Balbo, P. B., Wang, E. C. & Tsai, M. D. Kinetic mechanism of active site assembly and chemical catalysis of DNA polymerase  $\beta$ . *Biochemistry* **50**, 9865–9875 (2011).
47. Maximoff, S. N., Kamerlin, S. C. L. & Florián, J. DNA polymerase  $\lambda$  active site favors a mutagenic mispair between the enol form of deoxyguanosine triphosphate substrate and the keto form of thymidine template: a free energy perturbation study. *J. Phys. Chem. B* **121**, 7813–7822 (2017).
48. Freudenthal, B. D., Beard, W. A., Shock, D. D. & Wilson, S. H. Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157–168 (2013).
49. Freudenthal, B. D., Beard, W. A., Cuneo, M. J., Dyrkheeva, N. S. & Wilson, S. H. Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **22**, 924–931 (2015).

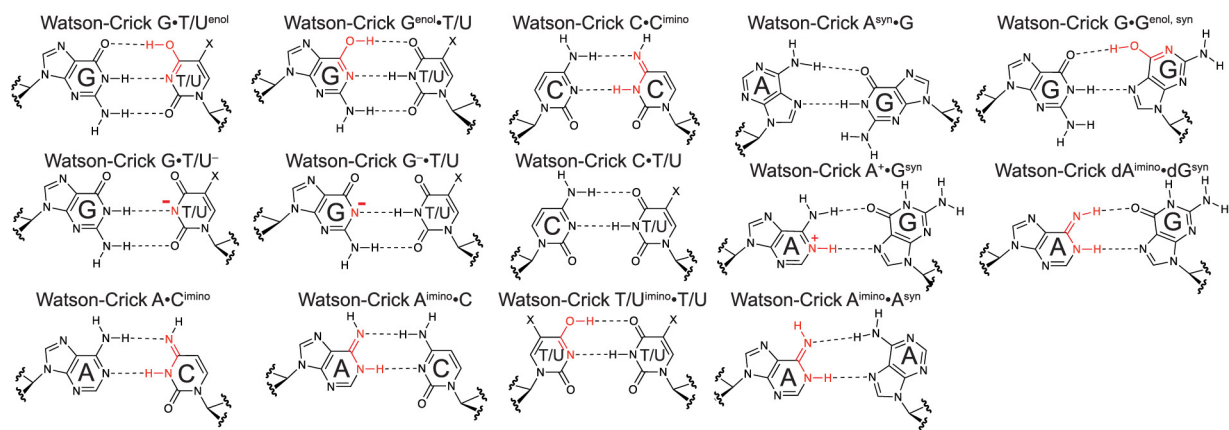
**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank members of the Al-Hashimi laboratory and T. Oas for discussions and input. We acknowledge technical support and resources from the Duke Magnetic Resonance Spectroscopy Center and the Duke Shared Cluster Resource. This work was supported by grants from the National Institutes of Health (NIH R01GM089846, P01GM0066275 and P50GM103297) and an Agilent Thought Leader Award to H.M.A., and a grant from the National Science Foundation (MCB-1716168) to Z.S. W.J.Z. was supported by a Pelotonia Graduate Fellowship from The Ohio State University Comprehensive Cancer Center.

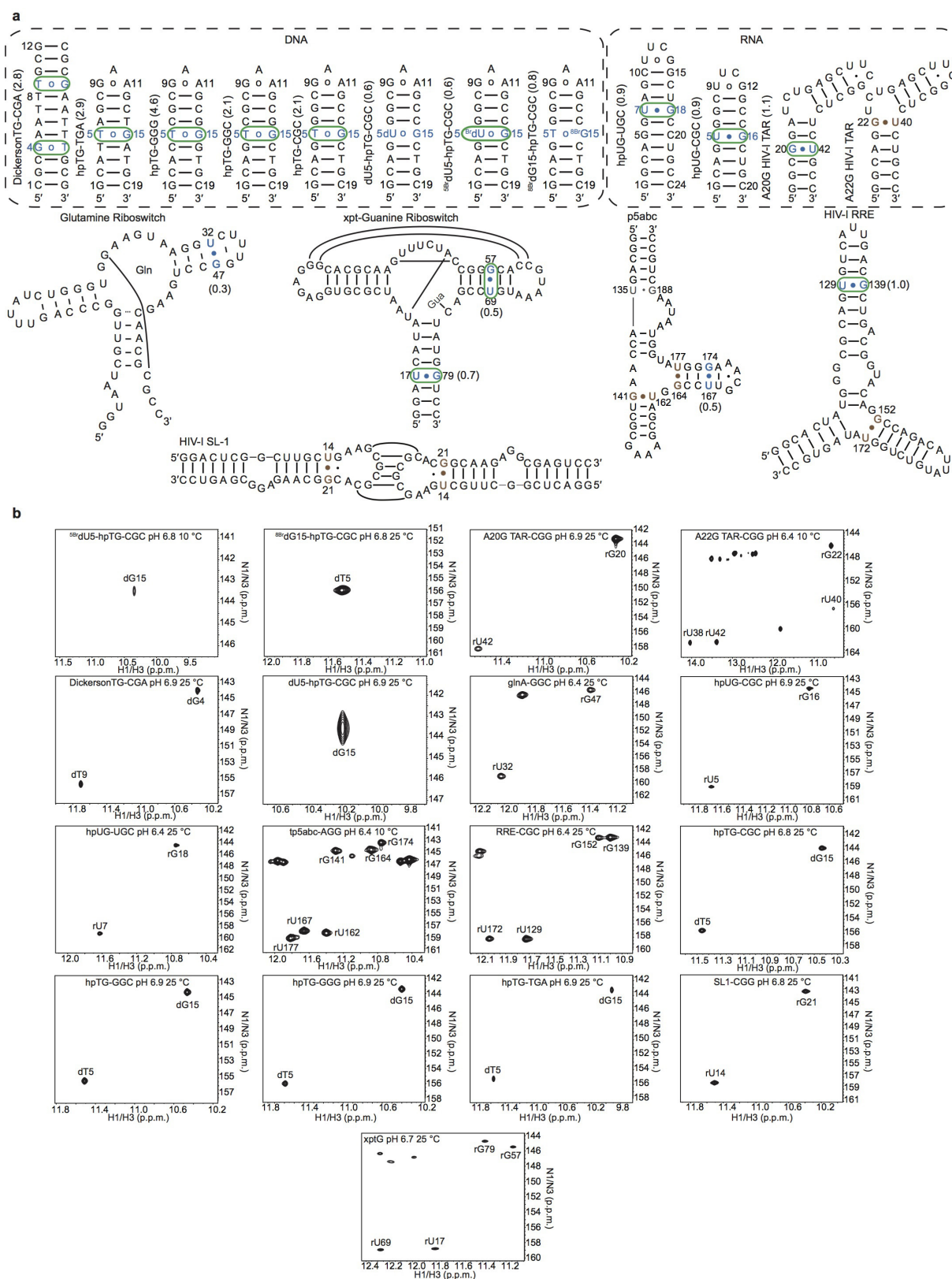
**Author Contributions** I.J.K., E.S.S., and H.M.A. conceived the NMR research and kinetic model. W.J.Z. and Z.S. conceived the kinetic experiments. I.J.K., E.S.S., W.J.Z., Z.S. and H.M.A. wrote the manuscript. I.J.K. synthesized all DNA constructs as well as the RNA hpUG-CGC, hpUG-CGU, and xptG riboswitch constructs, and collected and analysed all NMR relaxation dispersion data. E.S.S. and H.M.A. constructed and tested the kinetic models of misincorporation with input from Z.S. W.J.Z. performed all kinetic experiments and provided input on the kinetic model and simulations with help from Z.S. A.S. synthesized A20G and A22G HIV-I TAR constructs. Y.X. synthesized and assigned p5abc and glnA riboswitch constructs. C.-C.C. synthesized and assigned the HIV-I RRE construct. B.S. synthesized and assigned the HIV-I SL1 dimer complex.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to Z.S. ([suo.3@osu.edu](mailto:suo.3@osu.edu)) or H.M.A. ([hashim.al.hashimi@duke.edu](mailto:hashim.al.hashimi@duke.edu)).

**Reviewer Information** Nature thanks J. Essigmann, M. Goodman and E. Westhof for their contribution to the peer review of this work.



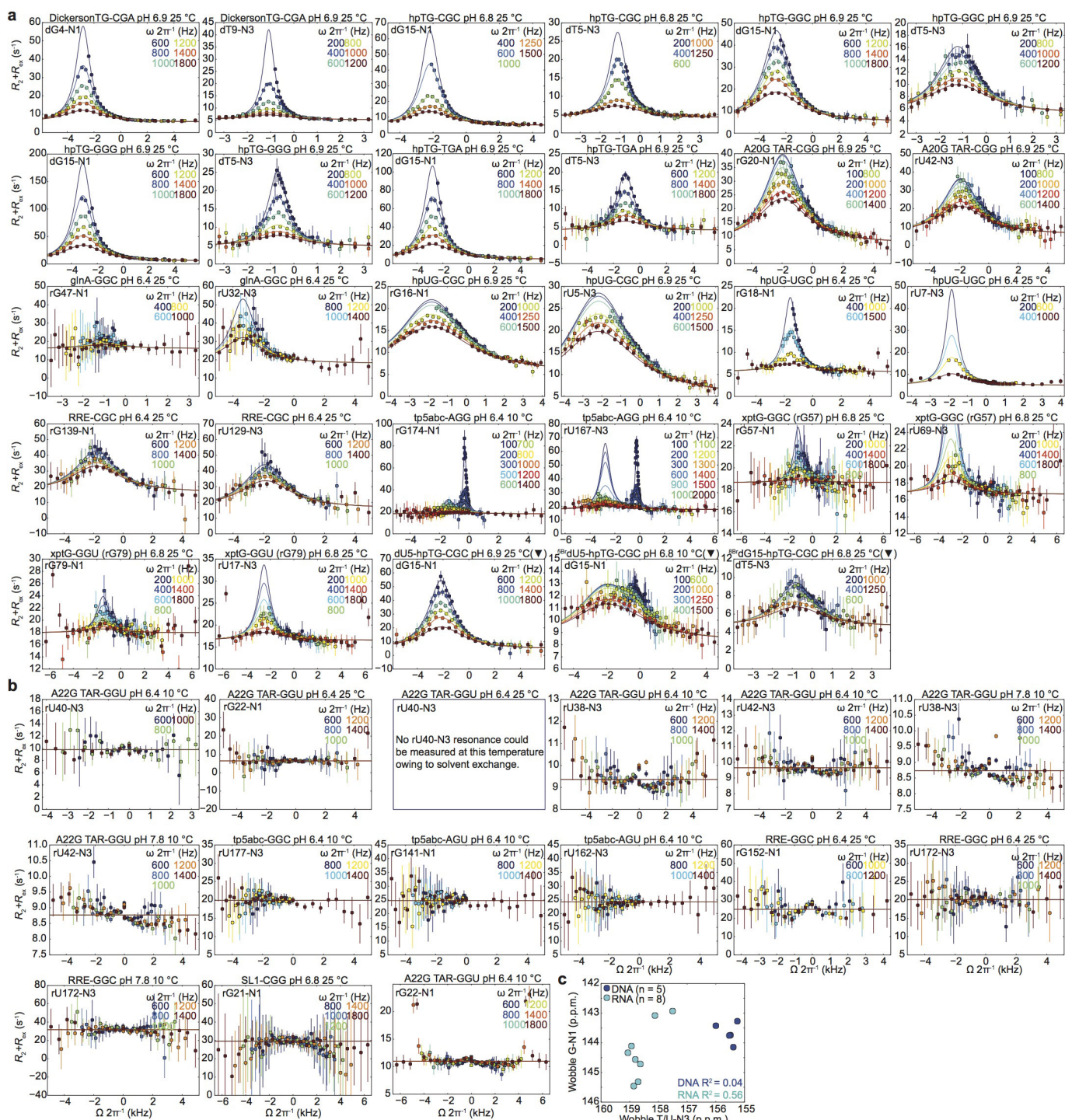
**Extended Data Figure 1 | Watson-Crick-like mismatches. a,** Watson-Crick-like mismatches stabilized by tautomeric and ionic base forms.



### Extended Data Figure 2 | DNA and RNA constructs used in this study.

**a**, Secondary structures of the various DNA and RNA constructs used in this study. G•T/U mismatches that show signs of chemical exchange directed towards tautomeric and/or anionic WC-like mismatches are highlighted in blue and green, respectively. G•T/U mismatches that show no evidence for WC-like relaxation dispersion are highlighted in brown. The value of  $K_t$  measured at near-neutral pH is shown next to

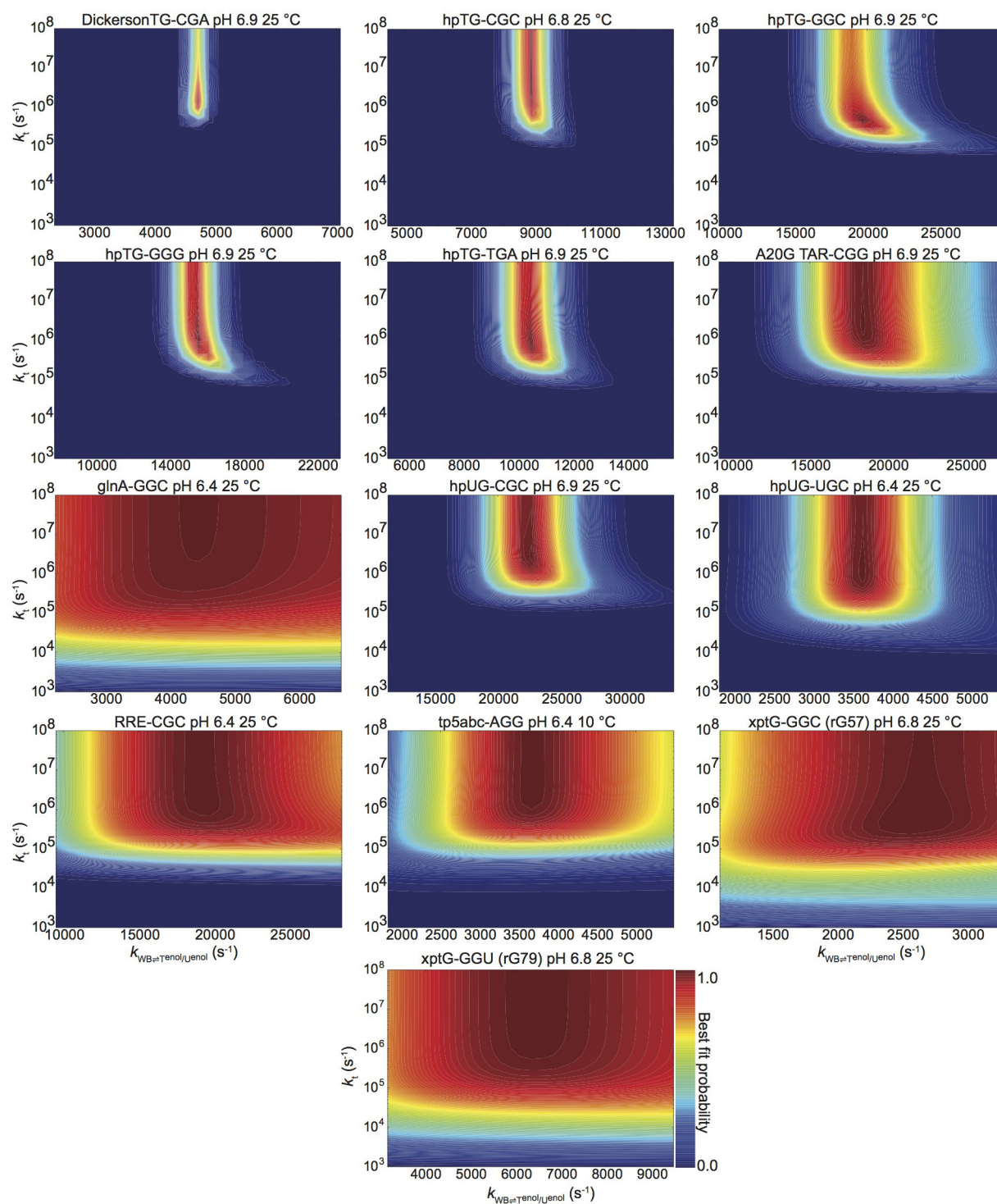
each mismatch. The DickersonTG-CGA, hpTG-CGC and hpUG-CGC sequences contexts were studied in a previous publication<sup>8</sup>. **b**, 2D [<sup>15</sup>N, <sup>1</sup>H] HSQC spectra of DNA and RNA constructs used in this study showing the imino resonances of G-N1/H1 and T/U-N3/H3 targeted for relaxation dispersion measurements. The spectrum shown for xptG was collected at pH 6.7 and 25 °C in potassium acetate buffer as described previously<sup>8</sup>.



**Extended Data Figure 3 | Relaxation dispersion profiles measured in DNA and RNA at near-neutral pH.** a,  $^{15}\text{N}$  G-N1 and T/U-N3 relaxation dispersion measured for G•T/U mismatches at pH 6.4–6.9 and 10–25°C showing wobble  $\Rightarrow$  tautomer exchange. Note that in addition to wobble  $\Rightarrow$  tautomer exchange, tp5abc undergoes an independent, slower exchange process involving a change in secondary structure that is described in detail elsewhere<sup>50</sup>. The trend lines represent B–M two- or three-state fits. Constructs containing a chemically modified base are

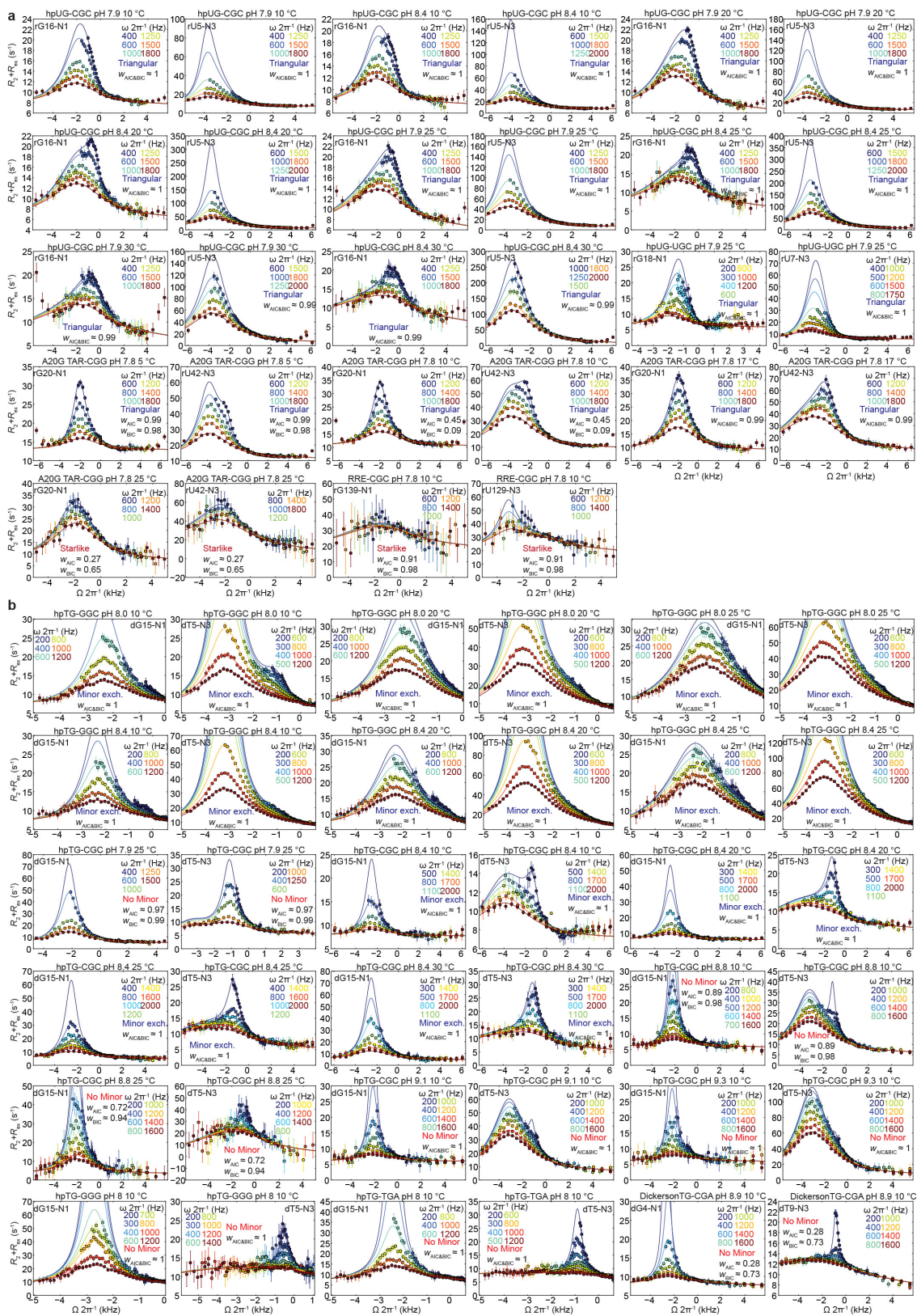
indicated by **▼**. **b**, The absence of  $^{15}\text{N}$  relaxation dispersion for rG•rU mismatches near bulges, apical loops or three-way junctions. Error bars reflect experimental uncertainty (one s.d., see Supplementary Methods). **c**, No correlation is observed between ground state (GS) wobble G-N1 and T/U-N3 chemical shifts for DNA ( $n=5$ ) or RNA ( $n=8$ ). Error bars in **a** and **b** reflect experimental uncertainty (one s.d., see Supplementary Methods).

50. Xue, Y. *et al.* Characterizing RNA excited states using NMR relaxation dispersion. *Methods Enzymol.* **558**, 39–73 (2015).



**Extended Data Figure 4 | Establishing lower limits for the rates of base pair tautomeric exchange.** Agreement between measured and predicted  $R_{1\rho}$  values (scaled  $\chi^2$  weight, equation (2) in Supplementary

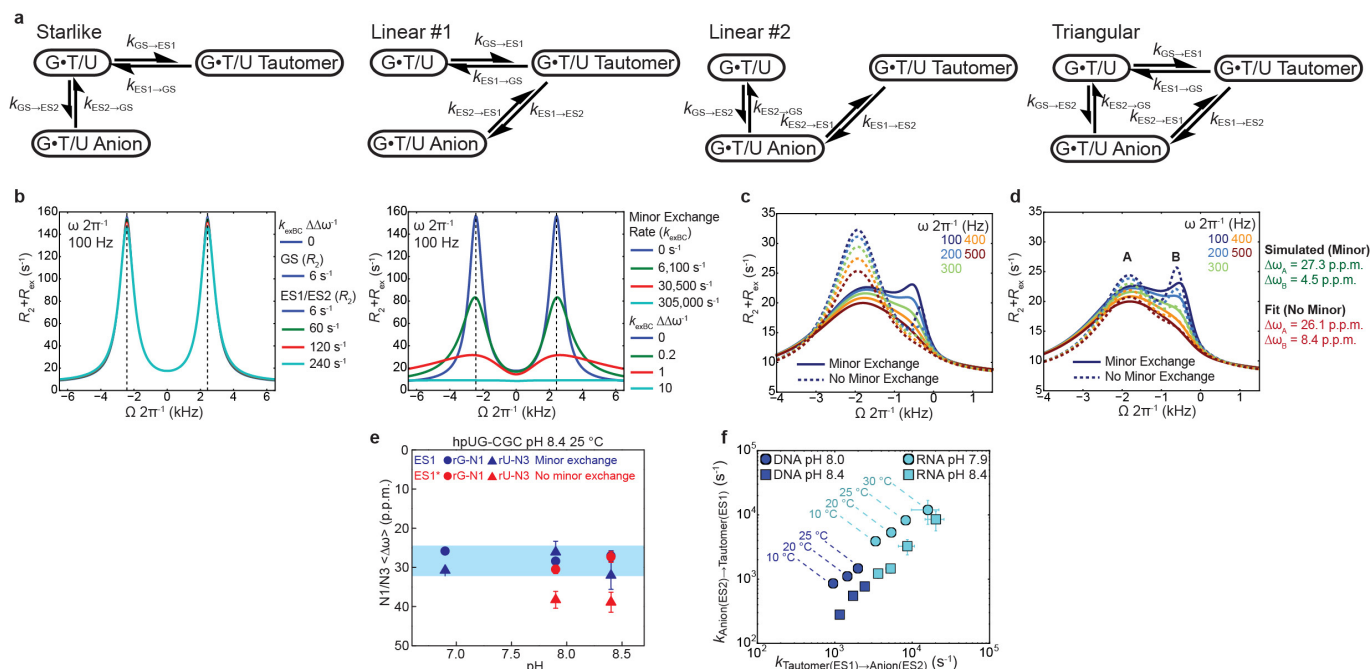
Methods) when varying the exchanges rates of wobble  $\rightleftharpoons G \bullet T^{\text{enol}}/U^{\text{enol}}$  ( $k_{WB \rightleftharpoons T^{\text{enol}}/U^{\text{enol}}}$ ) and  $G \bullet T^{\text{enol}}/U^{\text{enol}} \rightleftharpoons G^{\text{enol}} \bullet T/U$  ( $k_t$ ). See Supplementary Methods for additional details.



**Extended Data Figure 5 | Relaxation dispersion profiles measured in DNA and RNA at high pH.** **a, b,** B–M three-state fits of RNA (**a**) and DNA (**b**)  $^{15}\text{N}$  relaxation dispersion data for starlike and triangular topologies (as indicated within the plots). The relative statistical weights

$w_{\text{AIC}}$  and  $w_{\text{BIC}}$  (refs 51, 52) for each fit were used to select the model (representative starlike versus triangular, comparisons with linear models shown in Supplementary Table 6). Error bars reflect experimental uncertainty (one s.d., see Supplementary Methods).

51. Wagenmakers, E.-J. & Farrell, S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11**, 192–196 (2004).
52. Burnham, K. P. & Anderson, D. R. Multimodel inference. Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).

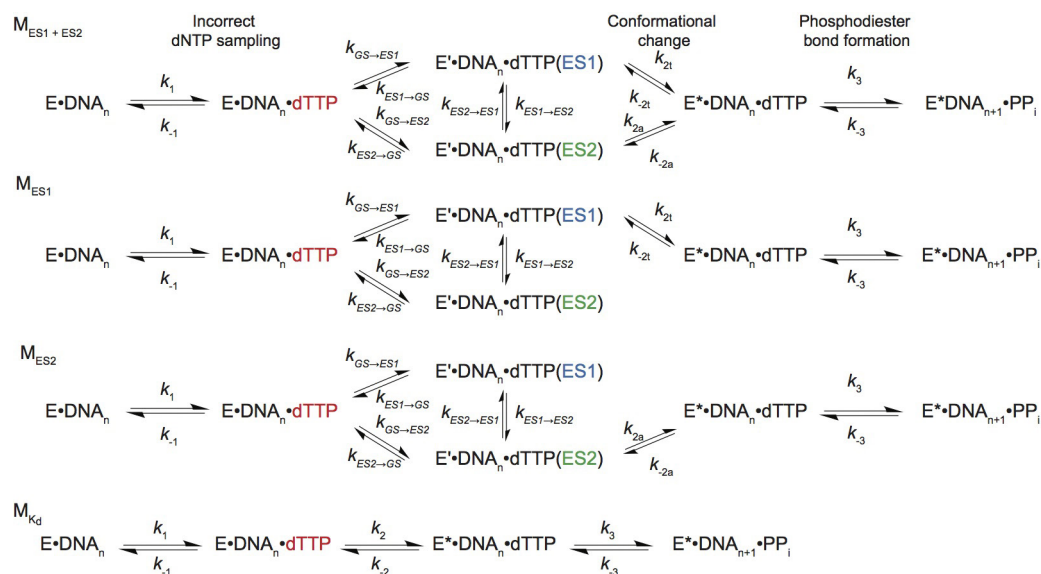


### Extended Data Figure 6 | Discerning minor exchange between

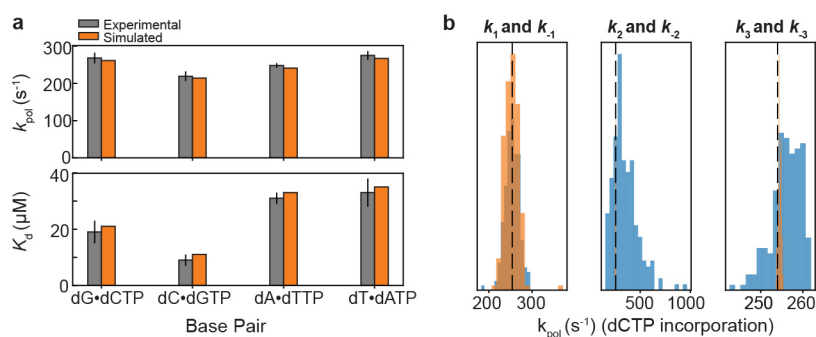
**WC-like tautomeric and anionic G•T/U mismatches** **a**, Topologies used to model chemical exchange. Individual rate constants are shown for each process of the different topologies. **b**, Left, B–M simulations showing that when  $R_{2(GS)} \neq R_{2(ES1)}$  and  $R_{2(GS)} \neq R_{2(ES2)}$ , no apparent peak asymmetry is observed. Right, B–M simulations showing that minor exchange between two ESs in a triangular topology induces asymmetry in the relaxation dispersion profiles and opposite changes in the apparent chemical shift for the two ESs. **c**, B–M simulations (solid lines) showing the fitted exchange parameters for hpUG-CGC at pH 8.4 and 10 °C (Supplementary Table 5) when including minor exchange in a triangular topology. For comparison, simulations using the same parameters without minor exchange ( $k_{ES1 \rightarrow ES2} = 0$  and  $k_{ES2 \rightarrow ES1} = 0$ ) are also shown (dashed lines). **d**, Dashed

lines denote

three-state B–M best fit to starlike topology ( $k_{ES1 \rightarrow ES2} = 0$  and  $k_{ES2 \rightarrow ES1} = 0$ ) to data simulated with triangular topology with minor exchange (solid lines). Shown to the right is the over- or underestimation of the true (green) versus fitted (red) ES chemical shifts when fitting relaxation dispersion profiles that have triangular topology with minor exchange to a starlike model that has no minor exchange. **e**, The ES1  $\langle \Delta\omega_{T-G-N1} \rangle$  and  $\langle \Delta\omega_{T-U-N3} \rangle$  values as a function of pH derived from the three-state B–M fit with triangular and starlike topology. **f**, Forward ( $k_{ES1 \rightarrow ES2}$ ) and reverse ( $k_{ES2 \rightarrow ES1}$ ) minor exchange rate constants for hpTG-GGC (pH 8 and 8.4) and hpUG-CGC (pH 7.9 and 8.4) as a function of temperature. Error bars in **e** and **f** reflect experimental uncertainty (one s.d., Supplementary Methods).

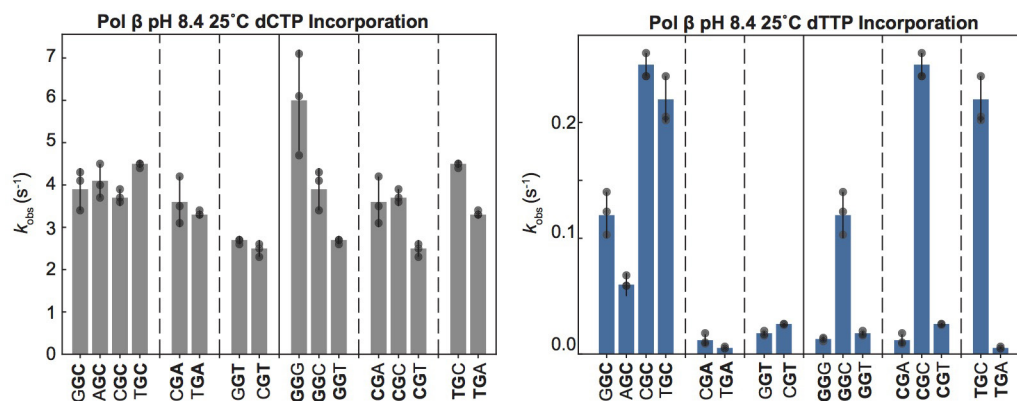


Extended Data Figure 7 | Kinetic mechanisms used to model misincorporation. Rate constants for each step are listed in Supplementary Table 7.



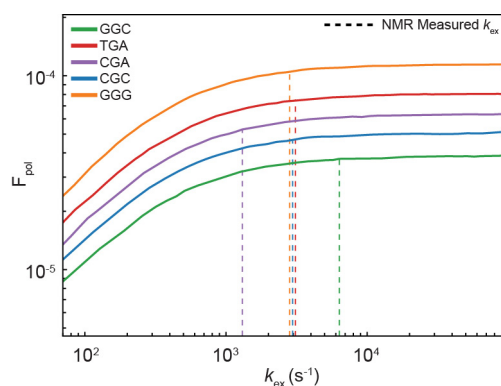
**Extended Data Figure 8 | Benchmarking kinetic simulations of misincorporation.** **a**, Comparison of  $k_{pol}$  and  $K_d$  values for correct incorporation measured experimentally for human DNA polymerase  $\epsilon$  with values computed on the basis of pre-steady state simulations using the microscopic rate constants provided in ref. 39. Error bars reflect fitting uncertainty as previously published<sup>39</sup>. **b**, Robustness of calculated

$k_{pol}$  values for human DNA polymerase  $\epsilon$  upon varying the rate constants (forward, blue; reverse, orange) for steps other than tautomerization or ionization by twofold ( $n = 200$  independent simulations in which rate constants were varied randomly by up to twofold). As expected, the only rate constant with a substantial effect on the reported  $k_{pol}$  values was the rate-limiting conformational change step  $k_2$  (middle).



**Extended Data Figure 9 |  $k_{obs}$  values measured for human DNA polymerase  $\beta$  insertion.** Incorporation of dCTP (dCTP•dG) is shown on the left, that of dTTP (dTTP•dG) on the right. pH 8.4, 25 °C, 100  $\mu$ M dNTP. DNA template sequence (5' to 3') is read from bottom

( $n + 1$  position) to top ( $n - 1$  position). Individual replicates ( $n = 3$  independent experiments) are indicated by grey circles. Bar height reflects average of replicates, and error bars reflect one s.d.



**Extended Data Figure 10 |  $F_{\text{pol}}$  is primarily governed by ES1 populations.** Simulated  $F_{\text{pol}}$  values as a function of scaling up or scaling down of the kinetic exchange rate for ES1 formation ( $k_{\text{ex}} = k_{\text{GS} \rightarrow \text{ES1}} + k_{\text{ES1} \rightarrow \text{GS}}$ ) without altering the ES1 population. Increasing  $k_{\text{ex}}$  beyond values measured experimentally in this study (green dotted line) minimally affects  $F_{\text{pol}}$ ; decreasing the  $k_{\text{ex}}$  within the range measured experimentally in this study (purple dotted line) also affects the value of  $F_{\text{pol}}$  only minimally. Much larger decreases in  $k_{\text{ex}}$  are required to significantly reduce the value of  $F_{\text{pol}}$ .

# Cryo-EM shows how dynactin recruits two dyneins for faster movement

Linus Urnavicius<sup>1\*</sup>, Clinton K. Lau<sup>1\*</sup>, Mohamed M. Elshenawy<sup>2</sup>, Edgar Morales-Rios<sup>3</sup>, Carina Motz<sup>1†</sup>, Ahmet Yildiz<sup>2,4</sup> & Andrew P. Carter<sup>1</sup>

**Dynein and its cofactor dynactin form a highly processive microtubule motor in the presence of an activating adaptor, such as BICD2. Different adaptors link dynein and dynactin to distinct cargoes. Here we use electron microscopy and single-molecule studies to show that adaptors can recruit a second dynein to dynactin. Whereas BICD2 is biased towards recruiting a single dynein, the adaptors BICDR1 and HOOK3 predominantly recruit two dyneins. We find that the shift towards a double dynein complex increases both the force and speed of the microtubule motor. Our 3.5 Å resolution cryo-electron microscopy reconstruction of a dynein tail–dynactin–BICDR1 complex reveals how dynactin can act as a scaffold to coordinate two dyneins side-by-side. Our work provides a structural basis for understanding how diverse adaptors recruit different numbers of dyneins and regulate the motile properties of the dynein–dynactin transport machine.**

Cytoplasmic dynein-1 (dynein) is the main transporter of cargoes towards the minus ends of microtubules in animal cells<sup>1</sup>. These cargoes move at a range of speeds<sup>2</sup> and vary in size from large organelles<sup>3</sup> to small individual proteins<sup>4</sup>. Dynein is activated to form a highly processive motor by binding its cofactor dynactin and a cargo adaptor, such as BICD2 (bicaudal D homologue 2)<sup>5,6</sup>. Dynein contains two motor domains joined by a tail region, whereas dynactin is built around a short actin-like filament, capped at its pointed and barbed ends and decorated with a shoulder<sup>6–9</sup>. A previous 8 Å resolution cryo-electron microscopy (cryo-EM) structure showed how a coiled coil in BICD2 recruits the tail of dynein to the filament of dynactin<sup>8</sup>. Other adaptors that activate dynein and link it to different cargoes have been identified<sup>5,10,11</sup>. These activating adaptors also contain long coiled coils; however, the sequence similarity between them is low<sup>12–15</sup> and it is unclear whether they engage dynein and/or dynactin in the same way as BICD2 does. There is also evidence that certain adaptors—such as BICDR1<sup>14</sup> (BICD related-1, also known as BICDL1) and HOOK3<sup>5,10,11</sup>—drive faster movement of dynein towards the minus ends of microtubules when compared with BICD2, although the mechanism underpinning this increased speed is not currently understood.

## Dynactin can recruit two dyneins

We determined the cryo-EM structures of two previously unsolved dynein–dynactin–adaptor complexes. BICDR1, like BICD2, binds RAB6 vesicles<sup>16</sup>, whereas HOOK3 links dynein and dynactin to early endosomes<sup>17,18</sup>. We determined 7 Å resolution maps of both the dynein tail–dynactin–BICDR1 complex (hereafter termed TDR) and the dynein tail–dynactin–HOOK3 complex (hereafter termed TDH), which we compare to the previously determined structure of the dynein tail–dynactin–BICD2 complex (hereafter termed TDB)<sup>8</sup> (Fig. 1a, Extended Data Fig. 1a–d, Extended Data Table 1).

The coiled coils of all three adaptors run along the length of the dynactin filament (Fig. 1a). However, in contrast to previous predictions<sup>13</sup>, each adaptor makes different interactions. BICD2 and BICDR1 diverge in their path and relative rotation (Fig. 1b). HOOK3 follows yet another route over the surface of dynactin (Fig. 1c). TDH also shows an extra

coiled-coil density near the pointed end of dynactin (Fig. 1c) and extra globular density towards its barbed end (Extended Data Fig. 1e, f). The identity of the second coiled coil is unclear, but the globular density probably corresponds to the N-terminal Hook domain, which is required for HOOK3 to activate dynein and dynactin<sup>11,19</sup>.

The most notable feature of TDR and TDH is the presence of two dynein tails (Fig. 1a). The tail of the first dynein (dynein-A) binds in an equivalent position to the dynein tail in TDB<sup>8</sup> and to the full-length dynein in the dynein–dynactin–BICD2 complex (hereafter termed DDB)<sup>9</sup>. The second dynein (dynein-B) binds next to dynein-A near the barbed end of dynactin.

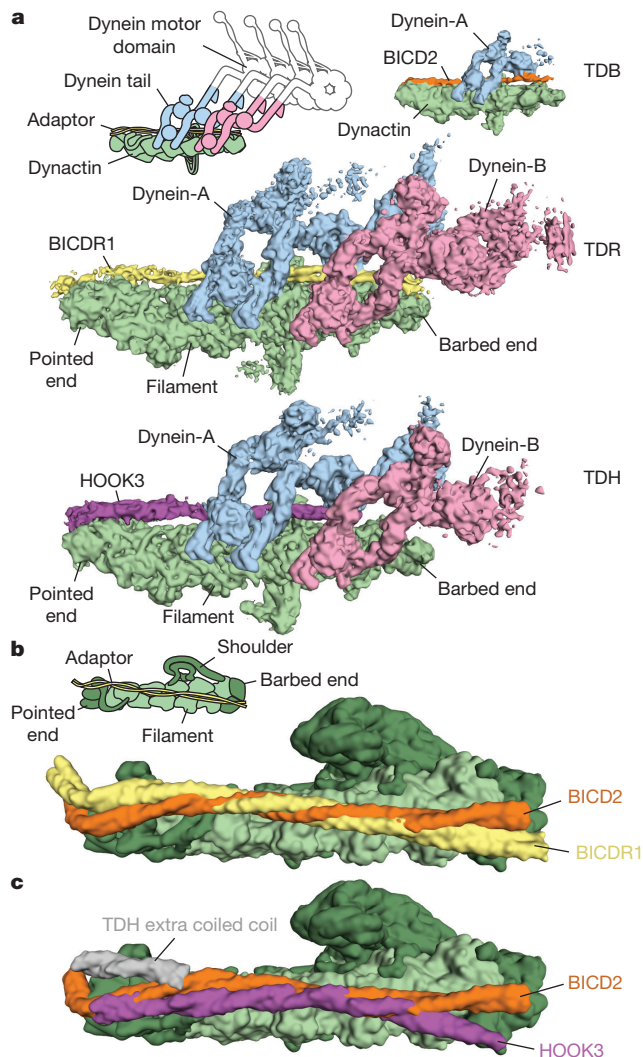
## Adaptors determine dynein recruitment

We determined whether BICD2, BICDR1 and HOOK3 recruited different numbers of dyneins in moving dynein–dynactin complexes. We mixed dyneins that had been labelled with tetramethylrhodamine (TMR) or Alexa Fluor 647, and used single-molecule fluorescence microscopy to measure the frequency at which the two dyes colocalized on microtubules (Fig. 2a, b). In the presence of dynactin and BICD2,  $13 \pm 1\%$  (s.e.m.) of processive complexes were labelled with both dyes; this was significantly higher ( $P < 0.0001$ ) than the colocalization observed for the dynein-only control ( $2.1 \pm 0.3\%$ ). Using BICDR1 or HOOK3 as an adaptor led to colocalization percentages of  $31 \pm 2\%$  and  $34 \pm 1\%$ , respectively (Fig. 2b). After correction for complexes that were double-labelled with the same colour, we estimate that 26% of BICD2 complexes contained two dyneins, compared to 61% for BICDR1 and 67% for HOOK3. We conclude that the majority of motile complexes that contain BICD2 have one dynein, whereas both BICDR1 and HOOK3 preferentially recruit two.

Although both this study and previous work<sup>6–9</sup> are consistent with BICD2 predominantly recruiting one dynein, its ability to recruit a second was unanticipated. To verify this observation, we applied a mixture of BICD2, dynein tail and dynactin onto grids for negative-stain electron microscopy analysis (Fig. 2c). In agreement with our single-molecule data, 3D classification of adaptor complexes showed that  $17 \pm 1\%$  of BICD2 complexes contained two dyneins;

<sup>1</sup>Medical Research Council Laboratory of Molecular Biology, Division of Structural Studies, Francis Crick Avenue, Cambridge CB2 0QH, UK. <sup>2</sup>Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA. <sup>3</sup>Department of Biochemistry, CINVESTAV, México D.F., Mexico. <sup>4</sup>Physics Department, University of California at Berkeley, Berkeley, California 94720, USA. <sup>†</sup>Present address: Department Chemie, Technische Universität München (TUM), Lichtenbergstrasse 4, 85747, Garching, Germany.

\*These authors contributed equally to this work.

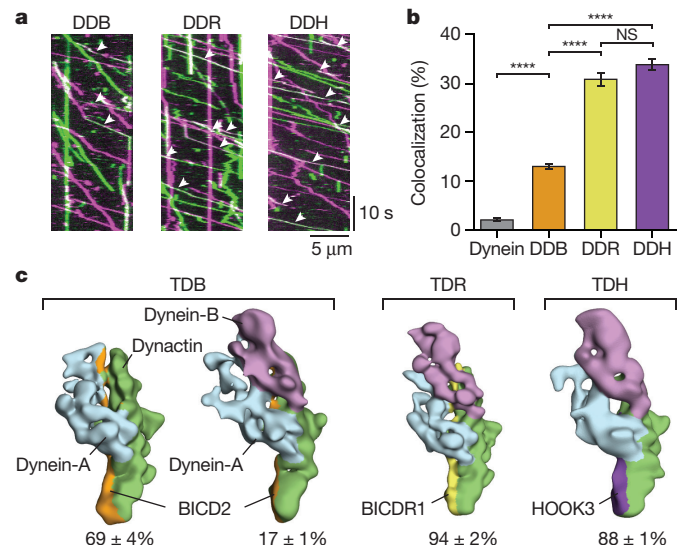


**Figure 1 | Dynactin can recruit two dyneins.** **a**, Sub-7 Å cryo-EM maps of TDR and TDH complexes, coloured according to their components. TDB complex (Electron Microscopy Data Bank (EMDB) code: EMD-2862) is included for comparison. **b**, Surface representation molecular models of BICDR1 and BICD2 on dynactin show the divergent paths of the coiled coils. **c**, Comparison of HOOK3 and BICD2 on dynactin.

69 ± 4% contained only one dynein, and the rest were ambiguous. The ability of BICD2 to bind two dyneins also agrees with a cryo-electron tomography reconstruction of microtubule-bound DDB<sup>20</sup>. Negative-stain electron microscopy of BICDR1 and HOOK3 complexes showed that 94 ± 2% and 88 ± 1% of these complexes, respectively, contained two dyneins (Fig. 2c). This suggests an even higher degree of second dynein recruitment than is indicated by our single-molecule data. Our data suggest that the number of dyneins bound to dynactin can be controlled by the identity of the adaptor.

## Two dyneins increase force and speed

We next investigated how the recruitment of different numbers of motors affects the motile properties of the dynein–dynactin complex. We used an optical trap to measure the stall force of DDB, dynein–dynactin–BICDR1 (hereafter termed DDR) and dynein–dynactin–HOOK3 (hereafter termed DDH) (Fig. 3a, b). Similar to our previous measurements<sup>21</sup>, the stall force of DDB is  $3.7 \pm 0.2$  pN, which is significantly lower ( $P < 0.0001$ ) than the stall force of the plus-end-directed motor kinesin-1 ( $5.7 \pm 0.2$  pN)<sup>22</sup>. By comparison, the stall force of DDR is  $6.5 \pm 0.3$  pN and that of DDH is  $4.9 \pm 0.2$  pN (Fig. 3b), which suggests that recruiting higher numbers of dyneins to dynactin increases force



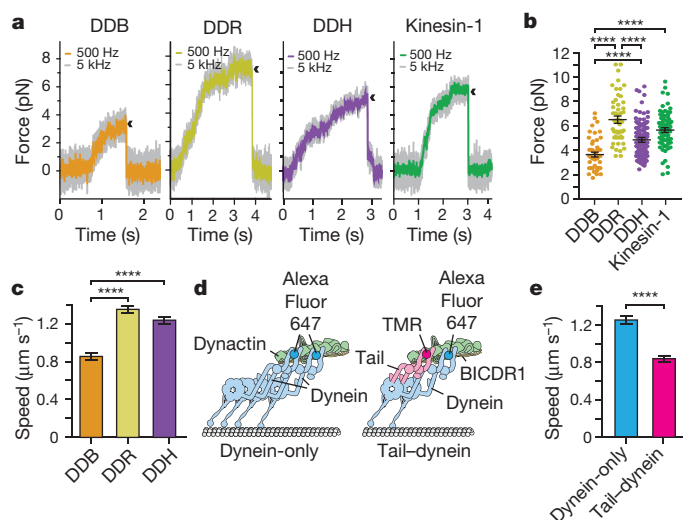
**Figure 2 | Different adaptors recruit different numbers of dyneins to dynactin.** **a**, Kymographs of DDB, DDR and DDH. Moving complexes contain TMR–dynein (magenta), Alexa Fluor 647–dynein (green) or both (white, white arrowheads). Experiment was independently repeated twice, with over three replicates per repeat. **b**, The mean percentage ± s.e.m. of complexes containing both TMR- and Alexa Fluor 647–dynein ( $n = 7,793$  (dynein),  $n = 3,092$  (DDB),  $n = 3,107$  (DDR),  $n = 3,990$  (DDH); ANOVA with Tukey's test; \*\*\*\* $P < 0.0001$ ; NS, not significant,  $P = 0.4010$ ). **c**, Representative negative-stain electron microscopy 3D classes of one- and two-dynein complexes for TDB, TDR and TDH. Mean ± s.e.m. fraction of particles in each class are shown. Ambiguous classes are not shown.

production. This agrees with previous reports that concluded—on the basis of using dyneins with beads—that dyneins can team up efficiently for maximum mechanical output<sup>23,24</sup>. The difference in stall force between the DDR and DDH dynein–dynactin complexes suggests that features other than motor number can also fine-tune force production.

The higher stall force of DDR also suggests that it competes more efficiently with kinesin than DDB does. This may explain why neuronal overexpression of BICDR1, but not of BICD2, counteracts kinesin-driven transport of RAB6 vesicles<sup>14</sup> and may be relevant to the role of BICDR1 in opposing anterograde movement in early neuronal differentiation<sup>16</sup>. The ability of some adaptors to recruit multiple dyneins could also contribute to the clustering and pairing of dynein motors required to transport large cargoes<sup>24,25</sup>.

We next explored whether the recruitment of more dyneins to dynactin had an effect on speed. Previous work on BICDR1 in cells<sup>14</sup> and HOOK3 *in vitro*<sup>5,10,11,19</sup> has shown that complexes containing these adaptors move faster than those containing BICD2. Our data raise the possibility that these faster speeds are due to an increased number of complexes with two dyneins. However, previous reports have suggested that, although artificially tethering dyneins increases run length, it has little or no effect on velocity<sup>26,27</sup>.

To determine whether motor number affects the movement speed of dynein–dynactin complexes, we first directly compared all three adaptors in our *in vitro* motility assay. As expected, the run lengths of DDR and DDH were longer than that of DDB (Extended Data Fig. 2a). Notably, the average velocities of DDR ( $1.35 \pm 0.04 \mu\text{m s}^{-1}$ ) and DDH ( $1.23 \pm 0.04 \mu\text{m s}^{-1}$ ) were significantly faster than that of DDB ( $0.86 \pm 0.04 \mu\text{m s}^{-1}$ ,  $P < 0.0001$ ) (Fig. 3c, Extended Data Fig. 2b, c). To investigate whether this difference in speed required the presence of two active dyneins, we mixed Alexa Fluor 647-labelled dynein with a TMR-labelled tail construct, BICDR1 and dynactin (Fig. 3d). For this experiment, we used a mutated full-length dynein that binds to dynactin as strongly as the dynein tail does<sup>8,9</sup> but moves at wild-type velocities (Extended Data Fig. 2b, d). We compared the speeds of



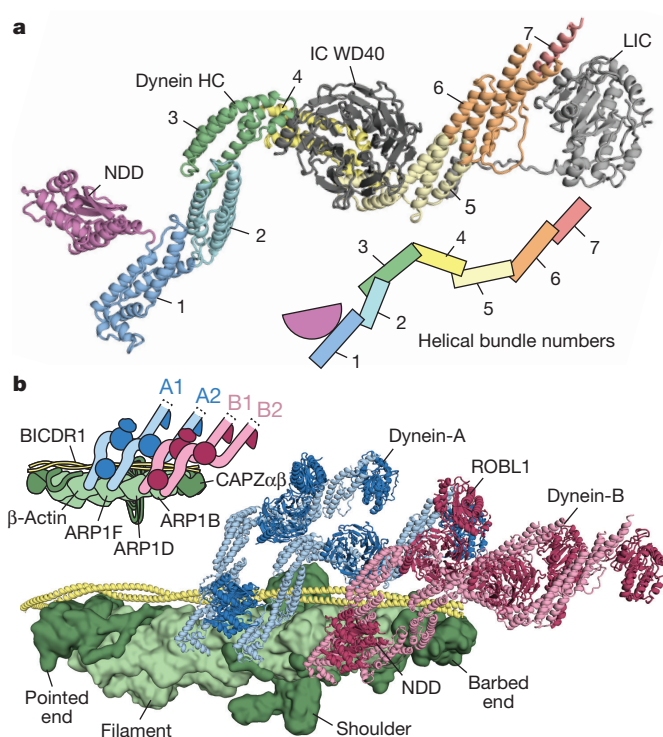
**Figure 3 | Two dyneins increase force and speed of dynein-dynactin.** **a**, Traces showing typical stalls of beads driven by single DDB, DDR, DDH or kinesin-1. Experiments were independently repeated four, six, five and four times for DDB, DDR, DDH and kinesin-1, respectively. Arrowheads denote motor detachment from the microtubule after the stall. **b**, Scatter plots showing stall force distributions ( $n = 54$  (DDB),  $n = 53$  (DDR),  $n = 118$  (DDH),  $n = 83$  (kinesin-1)). **c**, Mean speeds of DDB, DDR and DDH ( $n = 3,343$  (DDB),  $n = 3,162$  (DDR),  $n = 3,744$  (DDH)). **d**, Schematic depicting experimental design for TMR-tail-Alexa Fluor 647-dynein experiment. **e**, Dynein-only complexes move significantly faster than tail-dynein complexes ( $n = 1,004$  (dynein-only),  $n = 939$  (tail-dynein)). In **b**, **c** and **e**, horizontal lines represent mean  $\pm$  s.e.m., \*\*\*\* $P < 0.0001$  (ANOVA with Tukey's test for **b**, **c**; unpaired two-sided  $t$ -test for **e**).

moving complexes that contained only full-length dynein ('dynein-only') with those that contained one tail and one active dynein ('tail-dynein'). As expected, dynein-only complexes moved at a similar speed ( $1.25 \pm 0.04 \mu\text{m s}^{-1}$ , Fig. 3e) to DDR ( $1.22 \pm 0.05 \mu\text{m s}^{-1}$ , Extended Data Fig. 2d). However, tail-dynein complexes moved significantly more slowly ( $0.84 \pm 0.03 \mu\text{m s}^{-1}$ ,  $P < 0.0001$ , Fig. 3e, Extended Data Fig. 2e) than either DDR or dynein-only complexes. This suggests that the presence of a second dynein increases the velocity of dynein-dynactin complexes.

We propose that the increase in speed on the recruitment of two dyneins is linked to the way in which dynactin recruits them side-by-side (Fig. 1a). This may restrict the inherent sideways and backwards movements of the motor domains<sup>28</sup> and cause the complex to take a more direct and faster route along the microtubule. Other dynein regulators, such as LIS1 (otherwise known as PAFAH1B1), have been reported to increase the speed of dynein-dynactin complexes<sup>29,30</sup> and could act by increasing motor copy number. For LIS1, however, quantitative fluorescence measurements suggest that this is not the case<sup>29</sup>. The velocity of BICD2 complexes containing both fluorophores and therefore two dyneins ( $1.08 \pm 0.03 \mu\text{m s}^{-1}$ , Extended Data Fig. 2f) was significantly faster than the average DDB velocity ( $P < 0.0001$ ), but not as fast as DDR. This suggests that, in addition to recruiting two dyneins, certain adaptors also affect speed through small differences in how they recruit the motors to dynactin.

### The dynein-dynactin-BICDR1 structure

To investigate how dynactin recruits two dyneins, we collected data sufficient to determine the TDR structure to an overall resolution of 3.5 Å (Extended Data Fig. 3, Extended Data Table 1). To improve the tail density, we performed multiple rounds of particle signal subtraction, focused 3D classification and refinement on regions that moved as rigid blocks, which improved the definition of the blocks at each iteration (Extended Data Fig. 4). This produced a set of 3.4 Å maps that covered the entire length of the dynein tail (Extended Data Table 1).



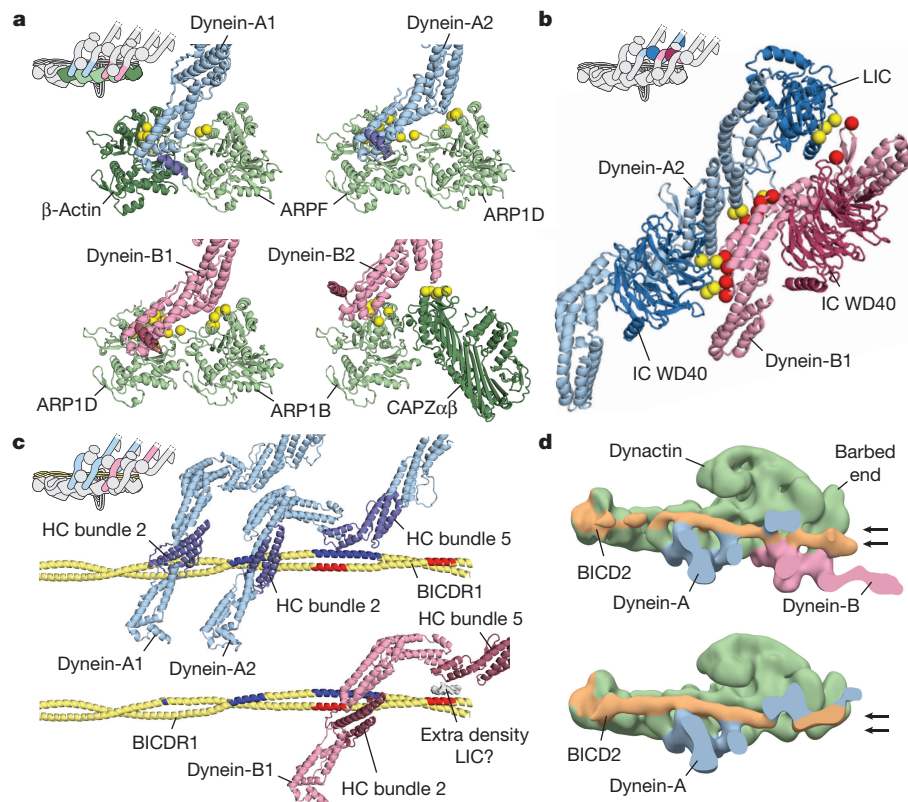
**Figure 4 | Structure of the dynein HC and architecture of the TDR complex.** **a**, Consensus molecular model of one dynein HC, complete with IC and LIC. HC is coloured according to helical bundle number. **b**, Assembled model of the TDR complex, showing the arrangement of dynein-A (cartoon) and dynein-B (cartoon) on BICDR1 (cartoon) and dynactin (surface). The HC NDD and dynein LC ROBL1 of dynein-B are labelled.

Previous low-resolution structures showed that the dynein tail comprises two heavy chains (HCs), which consist of a series of helical bundles held together by an N-terminal dimerization domain (NDD)<sup>8,9</sup>. Each HC binds one intermediate chain (IC) and one light intermediate chain (LIC)<sup>7,9,31</sup>. The IC N-terminal regions are held together by the dynein light chains, ROBL1 (otherwise known as DYNLRB1), LC8 (otherwise known as DYNLL1) and TCTEX1 (also known as DYNLT1)<sup>32,33</sup>. We used our high-resolution maps to build an atomic model of the dynein tail. We *de novo* traced helical bundles 1–6 of the HC and the WD40 domain of the IC (Fig. 4a, Extended Data Figs 5a, b, 6a, Extended Data Table 1). We also placed helices for part of helical bundle 7 and rebuilt homology models for the LIC<sup>31</sup> and ROBL1<sup>32</sup> (Fig. 4a, Extended Data Figs 5c, 6b, c). Our structure reveals that the IC WD40 domain makes extensive contacts with HC bundles 4 and 5, and that its central cavity is filled by a loop-helix from bundle 4 (Extended Data Fig. 6a). By contrast, the LIC globular domain interacts only with two helices from bundle 6. The tight binding of the LIC to the HC<sup>34</sup> is the result of its N- and C termini, which span out from the globular domain and form an integral part of HC bundles 5 and 7, respectively (Extended Data Figs 5c, 6b).

We assembled and refined a model of the whole TDR complex (Fig. 4b, Extended Data Table 1, Supplementary Video 1) into our 3.5 Å map. We used a previous dynactin structure<sup>8</sup> and a model of the BICDR1 coiled-coil region. For each dynein, we fit in two copies each of HC, IC and LIC, one ROBL1 dimer and a new 1.9 Å crystal structure of the human NDD (Extended Data Fig. 6d, e, Extended Data Table 2).

### Structural basis of two-dynein recruitment

Our TDR structure shows the two dyneins binding to grooves on the surface of dynactin that are formed by its  $\beta$ -actin subunit and the three actin-related protein 1 (ARP1, also known as ACTR1A) subunits ARP1F, ARP1D and ARP1B (Fig. 4b). The two dynein-A chains,



**Figure 5 | Interactions recruiting two dyneins to the TDR complex.** **a**, Dynein HC (dynein-A, blue; dynein-B, pink) interactions with dynactin subunits (green). Contact residues on dynactin are shown as yellow spheres. For each HC, helix  $\alpha 6$  is highlighted (dynein-A, dark blue; dynein-B, dark pink). **b**, A2 makes extensive interactions with B1. Interaction sites are shown as yellow and red spheres. **c**, Interactions of

dynein-A (top) and B1 (bottom) with BICDR1. Interaction sites marked in dark blue and red. Extra density from A2 LIC mediates the connection between B1 and BICDR1. **d**, Negative-stain electron microscopy reconstructions of DDB containing two dyneins (top) or dynein-A only (bottom), sliced to highlight BICD2. Arrows depict alternative positions of BICD2 at the barbed end of dynactin.

A1 and A2, and the first dynein-B chain, B1, all bind in a similar way, by making contacts to both sides of their respective grooves. Although the precise interactions overlap, they are all slightly different from one another (Fig. 5a). The second dynein-B chain, B2, binds between ARP1B and the barbed-end capping protein CAPZ $\beta$  (CAPZ $\beta$  is also known as CAPZB) and is rotated by 90° along its long axis, relative to the other dynein chains.

The two dyneins also make extensive interactions with one another. These consist of the IC WD40 domain of A2 binding the HC of B1; direct HC-to-HC interactions; and contacts between the A2 LIC and the HC and IC of B1 (Fig. 5b). These contacts are highly conserved across higher eukaryotes (Extended Data Fig. 7a). They contribute to a cascade of interactions (Supplementary Video 2) between the four dynein chains that include contacts between the IC WD40 domain of each chain and the neighbouring HC (Extended Data Fig. 7b). This network of connections stabilizes the binding of the second dynein and ensures all four HCs are held in a rigid orientation with respect to one another. This is likely to keep the dynein motor domains properly aligned and may be important for the increase in speed when two dyneins are recruited to the dynactin scaffold.

Our structure reveals the key role BICDR1 has in recruiting two dyneins to dynactin. Dynein-A binds the adaptor in three places: its A1 chain uses a single site on helical bundle 2, whereas its A2 chain binds via two sites (Fig. 5c). One of these sites also involves helical bundle 2, and the other uses helical bundle 5. Recruitment of dynein-B depends only on its B1 chain, which also uses sites on bundles 2 and 5. The first of these sites contacts the adaptor coiled coil in a position opposite the binding site of A2 (Fig. 5c). The second site does not directly contact the coiled coil, but instead touches density that packs against it (Fig. 5c, Extended Data Fig. 7c, d). Although the identity of this region

is uncertain, there is a weak density connecting it to the LIC, which suggests that it corresponds to the flexible LIC C terminus (Extended Data Fig. 5c, 7d). This region of LIC has previously been shown to interact with BICD2 and HOOK3<sup>19,31</sup>.

### Adaptor position controls dynein number

All three dynein–dynactin–adaptor complexes recruit dynein-A in a similar way, despite the differences in the positions of the adaptors themselves (Extended Data Fig. 8a). In TDR and TDH, dynein-B can contact the adaptor because the BICDR1 and HOOK3 N termini follow downward paths, stabilized by interactions with CAPZ $\beta$ . By contrast, in TDB no contact site for dynein-B is available because the adaptor is shifted upward towards the shoulder domain to contact ARP1A (Fig. 1b, Extended Data Fig. 8b).

We investigated which structural changes allow BICD2 to recruit a second dynein (Fig. 2). We combined our negative-stain electron microscopy TDB datasets (Fig. 2c) to determine structures of sufficient quality to resolve the position of the coiled coil. We found that TDB with two dynein tails has BICD2 in a lower position, similar to BICDR1 and HOOK3 and different from its position in single-dynein-bound TDB (Fig. 5d). Our data suggest that a switch in the position in the N terminus of the adaptor is sufficient to account for dynein-B recruitment.

In conclusion, we show that dynactin can act as a natural scaffold to line up two dyneins in close proximity. This arrangement results in a dynein–dynactin complex that moves faster and can produce larger forces when compared with complexes containing a single dynein. Our observations provide a mechanism by which cargo can control the output of the dynein–dynactin machine via the identity of its activating adaptor.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 31 August; accepted 18 December 2017.**

- Roberts, A. J., Kon, T., Knight, P. J., Sutoh, K. & Burgess, S. A. Functions and mechanics of dynein motor proteins. *Nat. Rev. Mol. Cell Biol.* **14**, 713–726 (2013).
- Klinman, E. & Holzbaur, E. L. F. Comparative analysis of axonal transport markers in primary mammalian neurons. *Methods Cell Biol.* **131**, 409–424 (2016).
- Burkhardt, J. K., Echeverri, C. J., Nilsson, T. & Vallee, R. B. Overexpression of the dynactin (p50) subunit of the dynactin complex disrupts dynein-dependent maintenance of membrane organelle distribution. *J. Cell Biol.* **139**, 469–484 (1997).
- Ben-Yaakov, K. *et al.* Axonal transcription factors signal retrogradely in lesioned peripheral nerve. *EMBO J.* **31**, 1350–1363 (2012).
- McKenney, R. J., Huynh, W., Tanenbaum, M. E., Bhabha, G. & Vale, R. D. Activation of cytoplasmic dynein motility by dynactin-cargo adaptor complexes. *Science* **345**, 337–341 (2014).
- Schlager, M. A., Hoang, H. T., Urnavicius, L., Bullock, S. L. & Carter, A. P. *In vitro* reconstitution of a highly processive recombinant human dynein complex. *EMBO J.* **33**, 1855–1868 (2014).
- Chowdhury, S., Ketcham, S. A., Schroer, T. A. & Lander, G. C. Structural organization of the dynein-dynactin complex bound to microtubules. *Nat. Struct. Mol. Biol.* **22**, 345–347 (2015).
- Urnavicius, L. *et al.* The structure of the dynactin complex and its interaction with dynein. *Science* **347**, 1441–1446 (2015).
- Zhang, K. *et al.* Cryo-EM reveals how human cytoplasmic dynein is auto-inhibited and activated. *Cell* **169**, 1303–1314.e18 (2017).
- Redwine, W. B. *et al.* The human cytoplasmic dynein interactome reveals novel activators of motility. *eLife* **6**, e28257 (2017).
- Olenick, M. A., Tokito, M., Boczkowska, M., Dominguez, R. & Holzbaur, E. L. Hook adaptors induce unidirectional processive motility by enhancing the dynein-dynactin interaction. *J. Biol. Chem.* **291**, 18239–18251 (2016).
- Gama, J. B. *et al.* Molecular mechanism of dynein recruitment to kinetochores by the Rod-Zw10–Zw12 complex and Spindly. *J. Cell Biol.* **216**, 943–960 (2017).
- Zheng, W. Probing the energetics of dynactin filament assembly and the binding of cargo adaptor proteins using molecular dynamics simulation and electrostatics-based structural modeling. *Biochemistry* **56**, 313–323 (2017).
- Schlager, M. A. *et al.* Bicaudal d family adaptor proteins control the velocity of Dynein-based movements. *Cell Reports* **8**, 1248–1256 (2014).
- Cianfrocco, M. A., DeSantis, M. E., Leschziner, A. E. & Reck-Peterson, S. L. Mechanism and regulation of cytoplasmic dynein. *Annu. Rev. Cell Dev. Biol.* **31**, 83–108 (2015).
- Schlager, M. A. *et al.* Pericentrosomal targeting of Rab6 secretory vesicles by Bicaudal-D-related protein 1 (BICDR-1) regulates neuritogenesis. *EMBO J.* **29**, 1637–1651 (2010).
- Bielska, E. *et al.* Hook is an adaptor that coordinates kinesin-3 and dynein cargo attachment on early endosomes. *J. Cell Biol.* **204**, 989–1007 (2014).
- Zhang, J., Qiu, R., Arst, H. N., Peñalva, M. A. & Xiang, X. HookA is a novel dynein-early endosome linker critical for cargo movement *in vivo*. *J. Cell Biol.* **204**, 1009–1026 (2014).
- Schroeder, C. M. & Vale, R. D. Assembly and activation of dynein-dynactin by the cargo adaptor protein Hook3. *J. Cell Biol.* **214**, 309–318 (2016).
- Grotjahn, D. A. *et al.* Cryo-electron tomography reveals that dynactin recruits a team of dyneins for processive motility. *Nat. Struct. Mol. Biol.* <http://dx.doi.org/10.1038/s41594-018-0027-7> (2018).
- Belyy, V. *et al.* The mammalian dynein-dynactin complex is a strong opponent to kinesin in a tug-of-war competition. *Nat. Cell Biol.* **18**, 1018–1024 (2016).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Rai, A. K., Rai, A., Ramaiya, A. J., Jha, R. & Mallik, R. Molecular adaptations allow dynein to generate large collective forces inside cells. *Cell* **152**, 172–182 (2013).
- Rai, A. *et al.* Dynein clusters into lipid microdomains on phagosomes to drive rapid transport toward lysosomes. *Cell* **164**, 722–734 (2016).
- Hendricks, A. G., Holzbaur, E. L. F. & Goldman, Y. E. Force measurements on cargoes in living cells reveal collective dynamics of microtubule motors. *Proc. Natl Acad. Sci. USA* **109**, 18447–18452 (2012).
- Driller-Colangelo, A. R., Chau, K. W. L., Morgan, J. M. & Derr, N. D. Cargo rigidity affects the sensitivity of dynein ensembles to individual motor pausing. *Cytoskeleton (Hoboken)* **73**, 693–702 (2016).
- Derr, N. D. *et al.* Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science* **338**, 662–665 (2012).
- Reck-Peterson, S. L. *et al.* Single-molecule analysis of dynein processivity and stepping behavior. *Cell* **126**, 335–348 (2006).
- Baumbach, J. *et al.* Lissencephaly-1 is a context-dependent regulator of the human dynein complex. *eLife* **6**, e21768 (2017).
- Gutierrez, P. A., Ackermann, B. E., Vershinin, M. & McKenney, R. J. Differential effects of the dynein-regulatory factor Lissencephaly-1 on processive dynein-dynactin motility. *J. Biol. Chem.* **292**, 12245–12255 (2017).
- Schroeder, C. M., Ostrem, J. M. L., Hertz, N. T. & Vale, R. D. A Ras-like domain in the light intermediate chain bridges the dynein motor to a cargo-binding region. *eLife* **3**, e03351 (2014).
- Hall, J., Song, Y., Karplus, P. A. & Barbar, E. The crystal structure of dynein intermediate chain–light chain roadblock complex gives new insights into dynein assembly. *J. Biol. Chem.* **285**, 22566–22575 (2010).
- Williams, J. C. *et al.* Structural and thermodynamic characterization of a cytoplasmic dynein light chain–intermediate chain complex. *Proc. Natl Acad. Sci. USA* **104**, 10028–10033 (2007).
- King, S. J., Bonilla, M., Rodgers, M. E. & Schroer, T. A. Subunit organization in cytoplasmic dynein subcomplexes. *Protein Sci.* **11**, 1239–1250 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Scheres, X. Bai, K. Vinothkumar and R. Leiro for cryo-EM advice; S. Chen, G. McMullan, C. Sava, G. Cannone, J. Grimmett and T. Darling for technical support; S. Bullock for SNAPf-dynein (1-1074-GST); M. Yu for crystallography support and the European Synchrotron Radiation Facility (beamline ID29) for data collection; T. Croll for model building; S. Bullock, L. Passmore, S. Lacey and H. Foster for manuscript comments; and G. Lander for discussions. This work was funded by Wellcome Trust (WT100387) and MRC grants (MC\_UP\_A025\_1011) to A.P.C.; and NIH (GM094522) and NSF (MCB-1055017, MCB-1617028) grants to A.Y.

**Author Contributions** L.U. performed all cryo-EM work on TDR, and C.K.L. performed all cryo-EM work on TDH. L.U., C.K.L., M.M.E. and A.P.C. performed single-molecule experiments. L.U. performed negative-stain electron microscopy. M.M.E. and A.Y. performed optical trapping. E.M.-R. determined the NDD structure. C.M. made dynein (1-1455). A.P.C., L.U. and C.K.L. built and refined the TDR model and prepared the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.P.C. ([cartera@mrc-lmb.cam.ac.uk](mailto:cartera@mrc-lmb.cam.ac.uk)).

## METHODS

**Cloning.** The following adaptors were codon-optimized for expression in Sf9 cells (Epoch Life Science): mouse BICDR1 (BICDL1), human HOOK3 residues 1–522 and mouse BICD2 residues 1–400. Adaptors were subcloned into pOmniBac- and pACEBac1-derived vectors for expression in Sf9 cells<sup>6</sup>. Tags were added for purification (a His<sub>6</sub>-ZZ tag with a TEV protease cleavage site (TEV), or a PreScission protease site (Psc) followed by a 2× strep affinity tag) or protein labelling (GFP or SNAPf (NEB)). The following constructs were generated: pOmniBac-His<sub>6</sub>-ZZ-TEV-SNAPf-BICD2<sup>1–400</sup>, pOmniBac-His<sub>6</sub>-ZZ-TEV-BICDR1, pOmniBac-His<sub>6</sub>-ZZ-TEV-SNAPf-BICDR1, pOmniBac-His<sub>6</sub>-ZZ-TEV-BICDR1-SNAPf, pOmniBac-His<sub>6</sub>-ZZ-TEV-BICDR1-GFP, pACEBac1-HOOK3<sup>1–522</sup>-SNAPf-Psc-2×strep and pACEBac1-HOOK3<sup>1–522</sup>-GFP-Psc-2×strep.

We generated a new dynein tail construct containing residues 1–1,455 of the human dynein HC. The fragment of the Sf9-codon-optimized *DYNC1H1* gene was cloned into a pACEBac1 vector contain an N-terminal His<sub>6</sub>-ZZ-TEV tag and fused to pDyn2 (containing genes for human IC2C, LIC2, TCTEX1, LC8 and ROBL1) as described<sup>6</sup>.

**Protein purification.** Dynactin was purified from pig brains using the large-scale SP sepharose protocol<sup>8</sup>. Wild-type human dynein and a mutant open dynein<sup>9</sup> were expressed and purified using baculovirus as described<sup>6</sup>. The two dynein tail constructs (HC<sup>1,455</sup> and SNAPf-HC<sup>1–1,074</sup>-GST) were purified as described<sup>8</sup>.

His<sub>6</sub>-ZZ-TEV-tagged adaptor constructs were purified as described<sup>6</sup>. C-terminal Psc-strep-tagged constructs were purified as follows: pellets from 500 ml of Sf9 cell culture were resuspended in 50 ml of lysis buffer (30 mM Hepes-KOH pH 7.2, 50 mM KAc, 2 mM MgAc, 1 mM EGTA, 10% (v/v) glycerol, 1 mM DTT) plus one complete-EDTA protease-inhibitor tablet (Roche) and 1 mM PMSF. Cells were dounced (30–40 strokes) and the lysate clarified in a Ti70 rotor (Beckman Coulter) at 30,000 r.c.f. for 20 min at 4°C before loading onto a 1-ml Streptap HP column (GE Healthcare) pre-equilibrated in lysis buffer. The column was washed with 20 column volumes of lysis buffer and bound protein eluted using 7 column volumes of lysis buffer plus 3 mM D-desthiobiotin. Protein-containing fractions were concentrated to approximately 5 mg ml<sup>−1</sup> in 30-kDa cut-off Amicon centrifugal filters (Merck Millipore) and gel-filtered using a Superose-6 10/300 column (GE Healthcare) pre-equilibrated in buffer containing 25 mM Hepes-KOH pH 7.2, 150 mM KCl, 1 mM MgCl<sub>2</sub>, 5 mM DTT.

A C-terminal GFP-tagged truncated human kinesin-1 (K560-GFP) was prepared as described<sup>35</sup>.

**N-terminal dimerization domain crystallization.** Residues 1–201 of *DYNC1H1* were expressed using a modified pRSET(A) plasmid<sup>36</sup>. Seleno-methionine-labelled NDD was expressed in a SoluBL21 *Escherichia coli* strain as described<sup>37</sup>. It was purified from 2 l of culture using a 5-ml HisTrap column (GE Healthcare). Fractions were pooled, concentrated in a 30-kDa Amicon and applied to a Superdex200 10/300 gel filtration column (GE Healthcare) equilibrated with buffer containing 50 mM Tris-HCl pH 7.4, 150 mM KAc, 10 mM β-mercaptoethanol, 2 mM MgAc, 1 mM EGTA, 10% (v/v) glycerol and inhibitor tablets (1 tablet: 100 ml, complete-EDTA free). The NDD peak was concentrated to 10 mg ml<sup>−1</sup>. For protein crystallization, 2 μl of protein was mixed with 2 μl precipitant (100 mM NaAc, pH 5.5, 10% (v/v) glycerol, 50 mM CaAc, 20% PEG 2,000 MME). Crystals were grown at 18°C by hanging drop for 48 h, collected with microloops (Mitegen), dipped into mother liquor containing an extra 15% (v/v) glycerol and flash-frozen in liquid nitrogen. Single-wavelength anomalous diffraction data were collected at ID29 beamline at ESRF, and integrated and scaled by the EDNA auto pipeline<sup>38</sup>. The structure was solved in PHENIX<sup>39</sup>, built in COOT<sup>40</sup> and refined using REFMAC<sup>41</sup>.

**Cryo-EM of TDR.** Cryo-EM grids for TDR (dynein tail (HC<sup>1,455</sup>)-dynactin-BICDR1) were prepared similarly to TDB<sup>8</sup>, though no cross-linker was used. Protein concentrations were chosen to give densely packed particles (approximately 100 per image). Micrographs were recorded using FEI Titan Krios equipped with Falcon II detector (300 kV, 32 frames, 2-s exposure, 1.34 Å per pixel, 52 e<sup>−</sup> per Å<sup>2</sup>) using automated data collection (EPU, FEI). Seven images were collected per hole (26,906 images in total, over 11 separate sessions). Drift correction was performed using MotionCor2<sup>42</sup> and contrast transfer function (CTF) parameters estimated using CTFFIND3<sub>130307</sub><sup>43</sup>. Subsequent processing used Relion v.2.0<sup>44</sup> unless otherwise stated.

Micrographs were first manually examined to remove images with a large amount of contamination, very low number of particles (<15), substantial uncorrected drift, a large astigmatism, extreme defocus values (<1 μm or >6 μm) and/or abnormal Fourier patterns. We selected 23,945 micrographs that had good signal to at least 8 Å. For the first dataset, a small set of particles was manually picked from 8× binned micrographs and subjected to reference-free 2D classification. A selection of 2D-class averages that represented a range of different size and shape particles present in micrographs (not just TDR complexes) was selected, centred using the Relion shift\_com function, low-pass filtered to 50 Å and used as references

for autopicking all binned micrographs using Relion v.1.4<sup>45</sup>. For optimal particle picking of other datasets, we used 2D classes obtained from multiple sessions. We also used a value of 1.1 for the ‘maximum standard deviation of the background noise’ setting to ensure we picked all good TDR particles with high contrast. This value resulted in substantial levels of contaminants but these were removed by extensive manual screening, as indicated below. Initially, particles with a high ‘autopick figure of merit’ values were screened to remove false positives. Then, the particles were cleaned by several cycles of 2D classification. At each cycle, the only particles that were discarded were those that were obviously dynein tail or contamination. In addition, particles with high ‘log likelihood contribution’ values were manually screened for remaining false positives and particles containing impurities with a very strong signal. After several cycles of 2D classification, particles corresponding to free dynactin in its dominant view were discarded. To do this, we first subjected particles from these classes to an additional round of 2D classification with the ‘image alignment’ setting turned off, in order to recover any TDR particles. Some of the dynactin classes of other-than-dominant view could not be easily distinguished from the TDR complex. Therefore, particles from all other dynactin-like classes were combined with all TDR particles and subjected to 3D refinement with the previous TDB structure<sup>8</sup> low-pass filtered to 60 Å as a reference. The output translational information from this 3D refinement was used to obtain more accurate coordinates of the particles in individual micrographs (script from R. Fernandez-Leiro). These coordinates were used to extract re-centred particles from 8-binned micrographs, which were manually cleaned as described above. After this cleaning, screened particles were subjected to another round of 3D refinement followed by 3D classification, to improve the separation of dynactin and TDR particles. In all the steps described above (picking, 2D classification, and 3D refinement and classification), the option to ‘ignore CTFs until first peak’ was turned on. Translational information from the 3D classification was used as above to extract re-centred TDR particles from unbinned micrographs. The 3D refinement using the unbinned particles from the first dataset yielded a 6.5 Å resolution map, and the 3D refinement using all 11 datasets yielded a 3.5 Å resolution map.

3D classification (see above) revealed movement of the dynein tails, which resulted in a lower resolution for these parts of the map. As a result, we conducted focused 3D classification and 3D refinement of the dynein tail as described<sup>46</sup>, except that we used multiple rounds of mask optimization. First, we generated overlapping binary masks covering the N- and C-terminal densities of all four dynein chains and used the particle subtraction function in Relion to subtract the density outside these regions from the raw TDR particles. Next, 3D refinement was used to align these ‘subtracted particles’ on the basis of the remaining density. Then we used 3D classification with no alignments to investigate which parts of the structure moved as a rigid block. We then made new masks around the rigid block and repeated all the steps of particle subtraction, 3D refinement and 3D classification. This process was repeated several times to obtain the optimal mask.

In the case of the N-terminal region of the tail, the optimal mask was used for a final round of particle subtraction and 3D refinement, which resulted in a 3.4 Å map. To further improve the densities for the IC WD40 domain, we performed local sub-volume averaging with Chimera 1.10<sup>47</sup>. Similarly, for the C-terminal region of the dynein tail we performed a round of particle subtraction and 3D refinement using the optimal mask. We then used 3D classification with no alignment to identify the most homogenous particles for different local regions. For HC residues 500–927, we performed a 3D classification using the whole of the optimal mask. For ROBL1 or LIC, we performed 3D classification using the local mass around their respective regions. In all cases, selected particles were refined using the whole optimal mask. All density maps were corrected for the modulation transfer function of the detector, and then sharpened by applying a negative *B* factor that was estimated either using automated procedures within Relion or by manually set parameters.

**Model building and refinement.** Building was performed in COOT and refinement in PHENIX. The dynein HC residues 201–710 from dynein-B1 and the IC from dynein-A2 were *de novo* built and refined into the ‘N-terminal tail’ map guided by maps generated by local sub-volume averaging, with improved density for flexible loops. HC residues 500–927 from dynein-A2 were built and refined into the ‘C-terminal tail’ map. A ‘LIC-mask’ map was used to model secondary structure elements for HC residues 927–1,057 and to rebuild a Phyre2<sup>48</sup> homology model for human LIC2 (both dynein-A2). A ‘Robl-mask’ map was used to rebuild a homology model for the ROBL1-IC-extended-N termini complex and identify its interactions with the dynein-A2 IC WD40 domain. The separately built components were combined to generate a consensus model for dynein-A2.

The structure of TDR was assembled and real-space refined into an overall TDR map that was not sharpened, and was filtered to 8 Å resolution. We used four copies of the HC-IC-LIC consensus model, two copies of the NDD, a model of dynactin<sup>8</sup> and a stretch of coiled coil for BICDR1. The combined model was fit

into the 3.5 Å overall map. Sections of dynactin, including the CAPZ $\alpha\beta$  dimer and the N terminus of subunit p50, were rebuilt. Corrections were made to secondary structure elements in the pointed end and shoulder domains. An approximate model for the BICDR1 coiled coil (residues 105–392) was generated by placing helices into density, and assigning registry on the basis of fitting residues Phe159 and Trp166 into the bulky density in the core of the coiled coil. Regions of the TDR model in weak density were set to zero occupancy for refinement into the overall 3.5 Å map. Segments of the final model were re-refined against the N-terminal and C-terminal maps (Extended Data Table 2).

**Cryo-EM of TDH.** To assemble the TDH complex, dynein tail, dynactin and HOOK3<sup>1–522</sup>–SNAPf were mixed in a 2:1:20 molar ratio and incubated on ice for 15 min. The sample was cross-linked to increase the amount of complex formed by the addition of 0.0125% (v/v) glutaraldehyde (Sigma-Aldrich) at room temperature for 15 min before quenching with 200 mM Tris pH 7.4 (final concentration). The sample was gel-filtered using a TSKgel G4000SW<sub>XL</sub> (TOSOH Bioscience) equilibrated in 25 mM Hepes-KOH pH 7.2, 150 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1 mM Mg.ATP, 5 mM DTT. The TDH complex was concentrated in a 100-kDa cut-off Amicon at 1,500 r.c.f. to 0.1–0.2 mg ml<sup>−1</sup>. Three microlitres of the TDH sample were applied to freshly glow-discharged Quantifoil R2/2 300-square-mesh copper grids covered with a thin carbon support. Samples were incubated on grids on a FEI Vitroblot IV for 45 s and blotted for 3–4.5 s at 100% humidity and 4 °C, and then plunged into liquid ethane.

Micrographs were recorded using automated data collection (FEI EPU) on an FEI Titan Krios with a FEI Falcon III detector in linear mode: 300 kV; 59 frames during a 1.5 s exposure; 1.42 Å per pixel; 45 e<sup>−</sup> per Å<sup>2</sup>; and 5 images per hole. Correction of inter-frame movement for each pixel and dose-weighting was performed using MotionCor2. CTF parameters were estimated using GCTF<sup>49</sup>. A reference set of 2D classes was generated using Relion v2.0 from a small set of particles picked by the EMAN2<sup>50</sup> Swarm boxing tool. Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) was used to pick particles from all micrographs (4× binned) using this reference set. Relion v2.0 was used for 2D classification to clean the autopicked particles. An 8.2 Å resolution TDB structure<sup>8</sup>, low-pass filtered to 60 Å, was used as an initial model for a first round of 3D refinement. The dataset was cleaned by 3D classification using the output from 3D refinement as a reference. The cleaned particles were re-extracted from unbinned micrographs and used for a final round of 3D refinement, yielding a 6.7 Å map.

**Negative stain electron microscopy analysis.** The dynein tail–dynactin–adaptor triple complexes were assembled by mixing 1.7 µl of 1 mg ml<sup>−1</sup> dynein tail (HC<sup>1,455</sup>), 1 µl of 1.35 mg ml<sup>−1</sup> dynactin and 2 µl of 1.3 mg ml<sup>−1</sup> cargo adaptor (BICD2<sup>1–400</sup>–GFP, HOOK3<sup>1–522</sup>–SNAPf and BICDR1–GFP). The samples were incubated on ice for 15 min before diluting eightfold for preparation of negative stain grids as described<sup>6</sup>. Two replicate samples were made on separate days. Between 200 and 400 micrographs of each sample were recorded using FEI EPU on a FEI Falcon II direct electron detector, fitted to a FEI F20 electron microscope operated at 200 kV: 1-s exposure; 2.08 Å per pixel; 20 e<sup>−</sup> per Å<sup>2</sup>; 0.8–1.2 µm underfocus. A small set of particles was picked from 4× binned micrographs using the EMAN2 Swarm boxing tool. Subsequent processing was done by Relion v2.0 unless otherwise stated. Particles were extracted and subjected to reference-free 2D classification. From two to five 2D-class averages of triple complex were centred using a shift\_com command, low-pass filtered to 50 Å resolution and used as references for automated particle picking within Relion v1.4 of all 4× binned micrographs. Autopicked particles were extracted, split in half and subjected to 2D classification as above. Ten 2D-class averages representing different particle orientations were selected and used for another round of autopicking. This third round of autopicking was used to obtain the optimal particle selection (fewest missed particles) and best centring.

The resulting particles were subjected to three rounds of 2D classification to identify good complex particles (particle numbers for TDB: 6,382, 7,648, 5,534, and 6,163; TDH: 3,430 and 3,958; TDR: 1,713, 1,861, 5,782 and 5,388). For the best 3D classification, these sets of particles were first subjected to 3D refinement using the cryo-EM structure of dynactin (EMDB code: EMD-2856), low-pass filtered to 60 Å, as a reference. 3D classification was carried out using the map from the 3D refinement, as well as the orientation and rotational assignment for the particles. 3D classification was carried out by setting the regularization parameter *T* to 2 and gradually adjusting image alignment sampling: 10 iterations with 15° angular sampling interval, offset search range set to 8 pixels and step to 2 pixels; 25 iterations with an angular sampling interval of 7.5°, offset search range set to 5 pixels and step to 1 pixel. The gradual reduction in sampling yielded the best classification. It was followed by 15 iterations with an angular sampling interval of 3.7°, offset search range set to 3 pixels and step to 0.5 pixels, which yielded 3D classes with density maps of sufficient quality to identify the number of dynein tails bound to dynactin. The fraction of complexes containing one or two dyneins was calculated from the number of particles assigned to each class.

To determine the path of BICD2, all particles of the complexes from both TDB (BICD2<sup>1–400</sup>–GFP) datasets were combined (42,823 particles). Binned particles from these complexes were subjected to 3D refinement as described above. These coordinate files were used to re-centre—as described above—and extract particles from unbinned micrographs, with CTF parameters that were determined using CTFIND3\_130307. Extracted particles were subjected to 3D refinement, followed by 3D classification. In both steps, the CTF correction was set to ignore CTFs until the first peak. Particles from 3D classes with dynein-A only, or with dynein-A and dynein-B, were separated and processed separately. Each set of particles was subjected to another round of 3D refinement and subsequent 3D classification (using 25 iterations: 7.5° angular sampling interval, offset search range of 5 pixels and step of 1 pixel) to obtain the best particles for each complex (13,278 and 14,070 particles for TDB with dynein-A only and with two dyneins, respectively). These were used in a final round of 3D refinement. Molecular models of dynactin, cargo adaptors and dynein tails were fitted into the density maps and used to colour different segments in Chimera.

**Single molecule assays.** SNAPf–dynein or SNAPf–tail complexes were labelled with TMR-Star or Alexa Fluor 647 (NEB) and purified separately, and the percentage labelling was quantified as previously described<sup>6</sup>. Single-molecule assays were conducted as previously described<sup>6</sup>, with slight modifications in order to optimize the number of moving complexes. Under these optimized conditions, DDB complexes moved faster than observed in previous publications<sup>5,6,10,11</sup>. All adaptor complexes were measured under identical conditions. The percentage of processive events was 56% for DDB, 76% for DDH and 75% for DDR.

Microtubules were made by mixing 26 µl of 5.2 mg ml<sup>−1</sup> unlabelled pig tubulin, 5 µl of 2 mg ml<sup>−1</sup> HiLyte Fluor 488 tubulin and 10 µl of 2 mg ml<sup>−1</sup> biotin tubulin (Cytoskeleton) in BRB80 buffer (80 mM PIPES pH 6.8, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT). The mixture was incubated on ice for 5 min before adding 41 µl of polymerization buffer (2× BRB80 buffer plus 20% (v/v) DMSO and 2 mM Mg.GTP). Microtubules were polymerized at 37 °C for 30–60 min. The sample was diluted with 200 µl of microtubule buffer (BRB80 plus 40 µM paclitaxel). Excess tubulin was removed by pelleting (20,238 r.c.f., 8.5 min; at room temperature). The microtubule pellet was washed with 200 µl of microtubule buffer and re-pelleted as above. The microtubule pellet was re-suspended in 200 µl of microtubule buffer and stored at room temperature, and covered from light for at least half a day (to a maximum of three days) before use. Microtubules were visualized (see below) and if the density was too low, or the free tubulin concentration was too high, they were re-pelleted and re-suspended in a smaller volume of microtubule buffer.

Motility chambers were made by applying two strips of double sided-tape (Tesa) approximately 8–10 mm apart on a glass slide and placing a cleaned coverslip<sup>6</sup> on top. The glass surface was functionalized with 0.4 mg ml<sup>−1</sup> biotinylated poly(L-lysine)-g-poly(ethylene-glycol) (SuSoS AG). The chamber was immediately washed with 40 µl of assay buffer (30 mM Hepes-KOH pH 7.2, 5 mM MgSO<sub>4</sub>, 1 mM EGTA, 1 mM DTT). Subsequently, 10 µl of 1 mg ml<sup>−1</sup> streptavidin (NEB) was flowed through and immediately washed with 40 µl of assay buffer. The chamber was then incubated with freshly diluted microtubules (typically 3 µl of microtubules and 10 µl of assay buffer). Microtubules were immediately washed out with assay buffer, followed by assay buffer supplemented with 1.25 mg ml<sup>−1</sup> α-casein (Sigma-Aldrich).

Dynein–dynactin–cargo–adaptor complexes were prepared by mixing 1 µl of 1 mg ml<sup>−1</sup> fluorescently labelled dynein, 1 µl of 1.35 mg ml<sup>−1</sup> dynactin and 2 µl of 1.3 mg ml<sup>−1</sup> cargo adaptor (SNAPf–BICD2<sup>1–400</sup>, HOOK3<sup>1–522</sup>–SNAPf and SNAPf–BICDR1). For experiments with SNAPf–tail, an additional 1 µl of 1 mg ml<sup>−1</sup> TMR-labelled tail was added. The complex was incubated on ice for 10–15 min, and then diluted with assay buffer to a final volume of 10 µl. One microlitre of complex solution was added to 19 µl of assay buffer containing 1.25 mg ml<sup>−1</sup> α-casein, an oxygen scavenging system (0.2 mg ml<sup>−1</sup> catalase (Calbiochem) and 1.5 mg ml<sup>−1</sup> glucose oxidase (Sigma-Aldrich)), 0.45% (w/v) glucose, 1% BME, 25 mM KCl and 5 mM Mg.ATP. Taxol was omitted from this buffer to reduce the non-specific background and non-moving events. The motility mix was flowed into the chamber, and washed with assay buffer supplemented with 1.25 mg ml<sup>−1</sup> α-casein for a second time. The sample was analysed immediately at 23 ± 1 °C with a total internal reflection fluorescence microscope, as previously described<sup>6</sup>. Colocalization and velocity were determined from the same datasets (4, 8, 9 and 7 chambers overall, for dynein-only, DDB, DDH and DDR, respectively). Data were collected on two different days. Tail and full-length dynein speed data were collected from three chambers.

Data were analysed using ImageJ<sup>51</sup>. Tif stacks were Z-projected to identify the paths of microtubules, segmented lines drawn along them and kymographs generated using the reslice function. Processive movements were defined as previously described<sup>6</sup>. Velocity was calculated using a pixel size of 105 nm and a frame rate of 236 ms per frame.

Colocalization data were collected using a DV2 beam splitter (Photometrics), which projected each channel (TMR and Alexa Fluor 647) onto a different half

of the image. Movie tif-stacks were split and kymographs generated for each channel. Kymographs were overlaid using the 'Colour:Merge' function to generate a composite image. The kymographs were manually scored for processive events that showed colocalization, followed by those events that only appeared in individual channels. Colocalization in the dynein-alone control chambers was determined for all microtubule binding events longer than 2 frames.

The fraction of complexes containing two dyneins,  $d$ , was calculated from the fraction of total events with signal for TMR-only ( $R_{\text{obs}}$ ), Alexa Fluor 647-only ( $G_{\text{obs}}$ ) and colocalized signal ( $Y_{\text{obs}}$ ). Colocalized events represent dynein–dynactin–adaptor complexes with two dyneins, and single-coloured events represent a mixture of both one- and two-dynein complexes. We can express this using the following three equations:

$$R_{\text{obs}} = (s \times r) + (d \times r^2)$$

$$G_{\text{obs}} = (s \times r) + (d \times g^2)$$

$$Y_{\text{obs}} = d \times 2(r \times g)$$

Where  $s$  is the fraction of complexes that contain one dynein,  $r$  is the fraction of TMR-labelled dynein and  $g$  is the fraction of Alexa Fluor 647-labelled dynein flowed into the chamber. These equations hold at high labelling efficiency—our dynein was >94% labelled for a dynein monomer, or >99.64% per dimer—where dynein can be labelled by either TMR or Alexa Fluor 647 ( $r + g = 1$ ). We can therefore solve these equations for  $d$  without knowing  $r$  or  $g$ .

**Stall force measurements.** Eight-hundred-nanometre carboxy latex beads (Life Technologies) were functionalized with anti-GFP antibodies<sup>21</sup>. Dynein, dynactin and cargo adaptor (BICD2<sup>1–400</sup>–GFP, HOOK3<sup>1–522</sup>–GFP and BICDR1–GFP) were mixed at a 1:5:20 molar ratio and incubated for 10 min on ice in dynein motility buffer (DMB: 30 mM HEPES pH 7.0, 5 mM MgSO<sub>4</sub>, 1 mM EGTA) supplemented with 1 mM tris(2-carboxyethyl)phosphine (TCEP) and 1 mg ml<sup>−1</sup> BSA. The diluted mixture was incubated with the anti-GFP-beads for 10 min on ice. Cy5-labelled axonemes<sup>21</sup> were introduced to the sample flow chamber, which was washed with 40 μl of DMB plus 1 mM TCEP and 500 μg ml<sup>−1</sup> casein. The protein–bead mixture was introduced to the chamber in imaging buffer (DMB with 1 mM TCEP, 500 μg ml<sup>−1</sup> casein, 2.5 mM protocatechuic acid, 35 μg ml<sup>−1</sup> protocatechuate-3,4-dioxygenase (PCD), 2 mM Mg.ATP). Kinesin was diluted in BRB80 supplemented with 1.5 mg ml<sup>−1</sup> casein and 2 mM DTT and mixed with anti-GFP beads for 10 min on ice. Kinesin-coated beads were introduced to the sample chamber in motility imaging solution (BRB80 supplemented with 2 mM DTT and 1.5 mg ml<sup>−1</sup> casein, 2.5 mM protocatechuic acid, 35 μg ml<sup>−1</sup> PCD, 2 mM Mg.ATP). For each experiment, the protein concentration was adjusted until less than 30% of the tested beads exhibited any binding and motility when brought in contact with axonemes. This ensured that 95% of the beads were driven by a single processive motor complex.

Trapping experiments were performed on a custom-built fully automated optical trap microscope setup<sup>21</sup>. To generate stall force histograms, position data from trap recordings (5 kHz) were down-sampled to 500 Hz and then manually screened and selected. To qualify as a stall event, the position trace had to reach a plateau and remain stationary for at least 100 ms before full release from the microtubule (defined as a rapid greater-than-2 ms jump of at least 50 nm). For DDB, we measured  $n = 54$  stalls from 14 beads in 4 independent experiments; for DDH,  $n = 118$  stalls from 25 beads in 6 independent experiments; for DDR,  $n = 53$  stalls from 17 beads in 5 independent experiments; and for kinesin,  $n = 83$  stalls from 18 beads in 4 independent experiments.

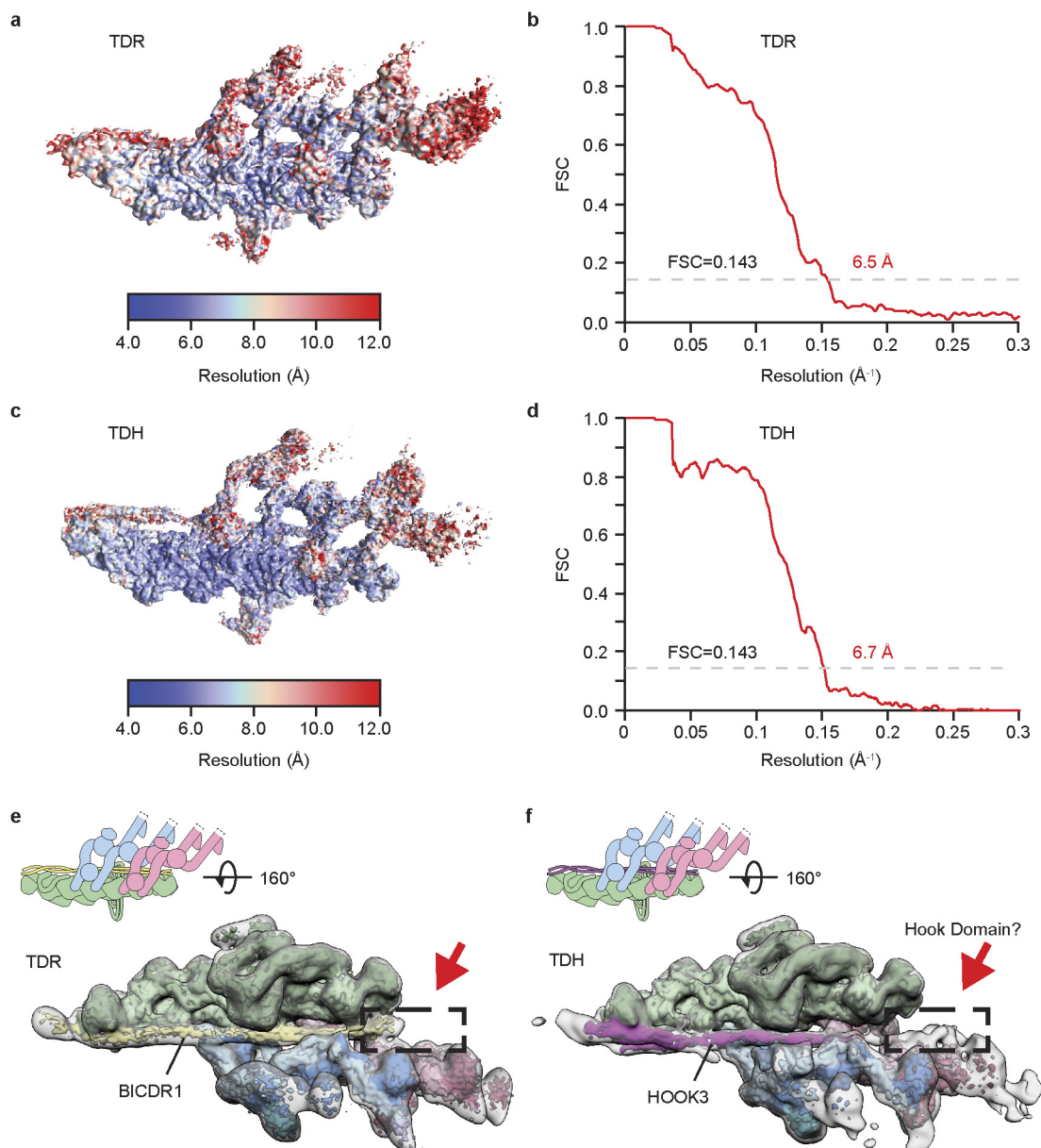
**Statistical analysis.** Statistical analysis was performed using GraphPad Prism 7 (GraphPad). The statistical significance of difference in mean values was calculated

using one-way ANOVA with Tukey's multiple comparisons test, or an unpaired  $t$ -test, as indicated for each experiment. No statistical methods were used to predetermine sample size. Velocity and colocalization data were randomized before analysis; other experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

**Code availability.** Code used is available from A.P.C.

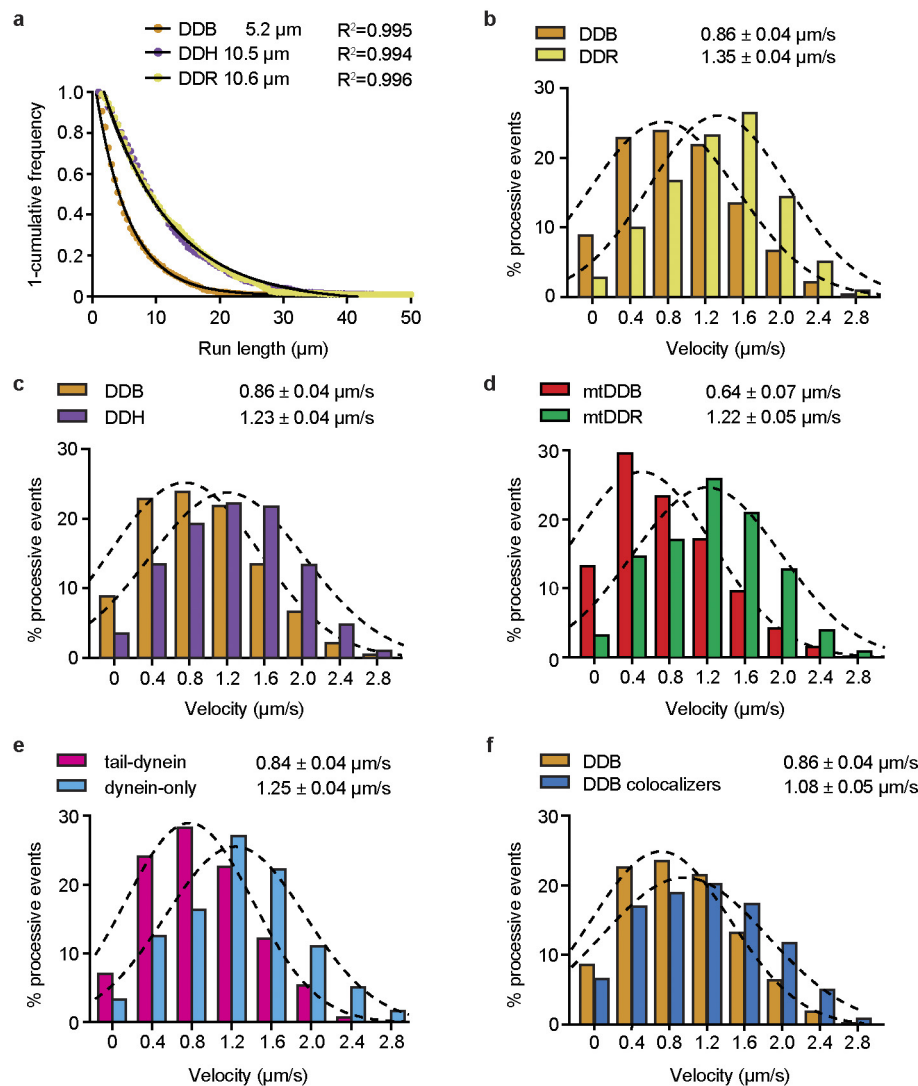
**Data availability.** Cryo-EM maps have been deposited in the EMDB under accession codes EMD-4168 (whole TDR complex), EMD-4169 (N-terminal tail), EMD-4170 (C-terminal tail/HCC terminus), EMD-4171 (LIC region), EMD-4172 (ROBL1 region) and EMD-4177 (TDH complex). Coordinates are available from the RCSB Protein Data Bank under accession codes 6F1T (whole TDR complex), 6F1U (N-terminal tail), 6F1V (C-terminal tail), 6F1Y (LIC region), 6F1Z (ROBL1 region), 6F38 (TDH complex) and 6F3A (TDB complex). Raw data are available from A.P.C.

35. Tomishige, M. & Vale, R. D. Controlling kinesin by reversible disulfide cross-linking. Identifying the motility-producing conformational change. *J. Cell Biol.* **151**, 1081–1092 (2000).
36. Dodd, R. B. *et al.* Solution structure of the Kaposi's sarcoma-associated herpesvirus K3 N-terminal domain reveals a novel E2-binding C4HC3-type RING domain. *J. Biol. Chem.* **279**, 53840–53847 (2004).
37. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
38. Incardona, M. F. *et al.* EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).
39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
41. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
42. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
43. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
44. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
45. Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. W. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**, 1743–1753 (2015).
46. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
47. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
48. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
49. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
50. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
51. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
52. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
53. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
54. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).



**Extended Data Figure 1 | Single particle cryo-EM analysis of TDR and TDH.** **a**, Cryo-EM reconstruction of the TDR complex analysed by ResMap<sup>52</sup> and showing resolution distribution from 4 to 12 Å. **b**, The gold-standard Fourier shell correlation (FSC) curve of the 6.5 Å TDR map. **c**, Cryo-EM reconstruction of the TDH complex, showing resolution distribution from 4 to 12 Å. **d**, The gold-standard FSC curve of the 6.7 Å TDH map. **e**, Cryo-EM density for TDR low-pass filtered to

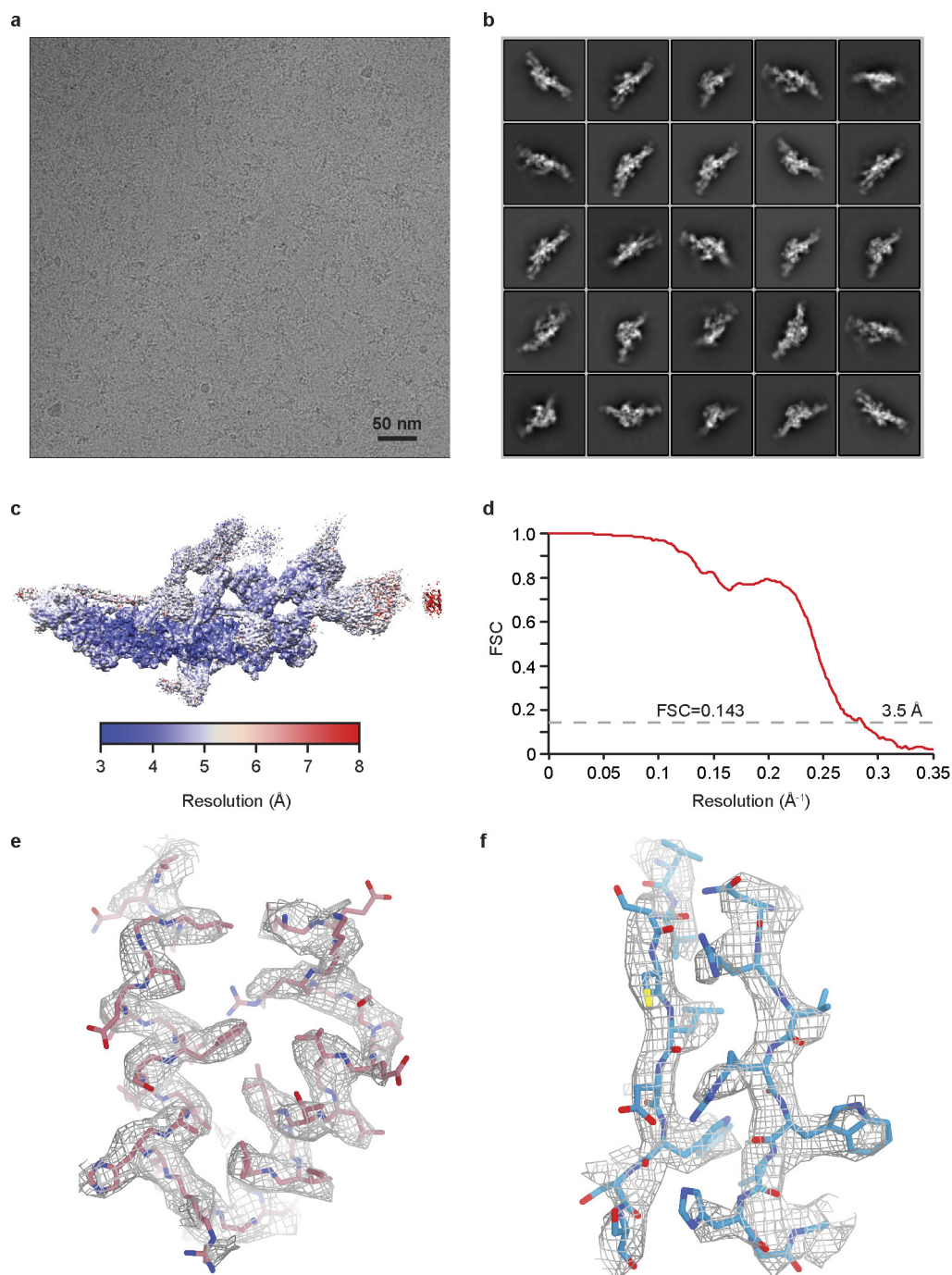
6.7 Å resolution (coloured according to cartoon) and to 20 Å (transparent outline). Density at the N terminus of BICDR1 is boxed. **f**, Cryo-EM density for TDH low-pass filtered to 6.7 Å (coloured according to cartoon) and to 20 Å (transparent outline) showing the putative Hook domain, an extension of the HOOK3 coiled coil ending in extra density near dynein-B (dashed box).



### Extended Data Figure 2 | Single-molecule assay speed distributions.

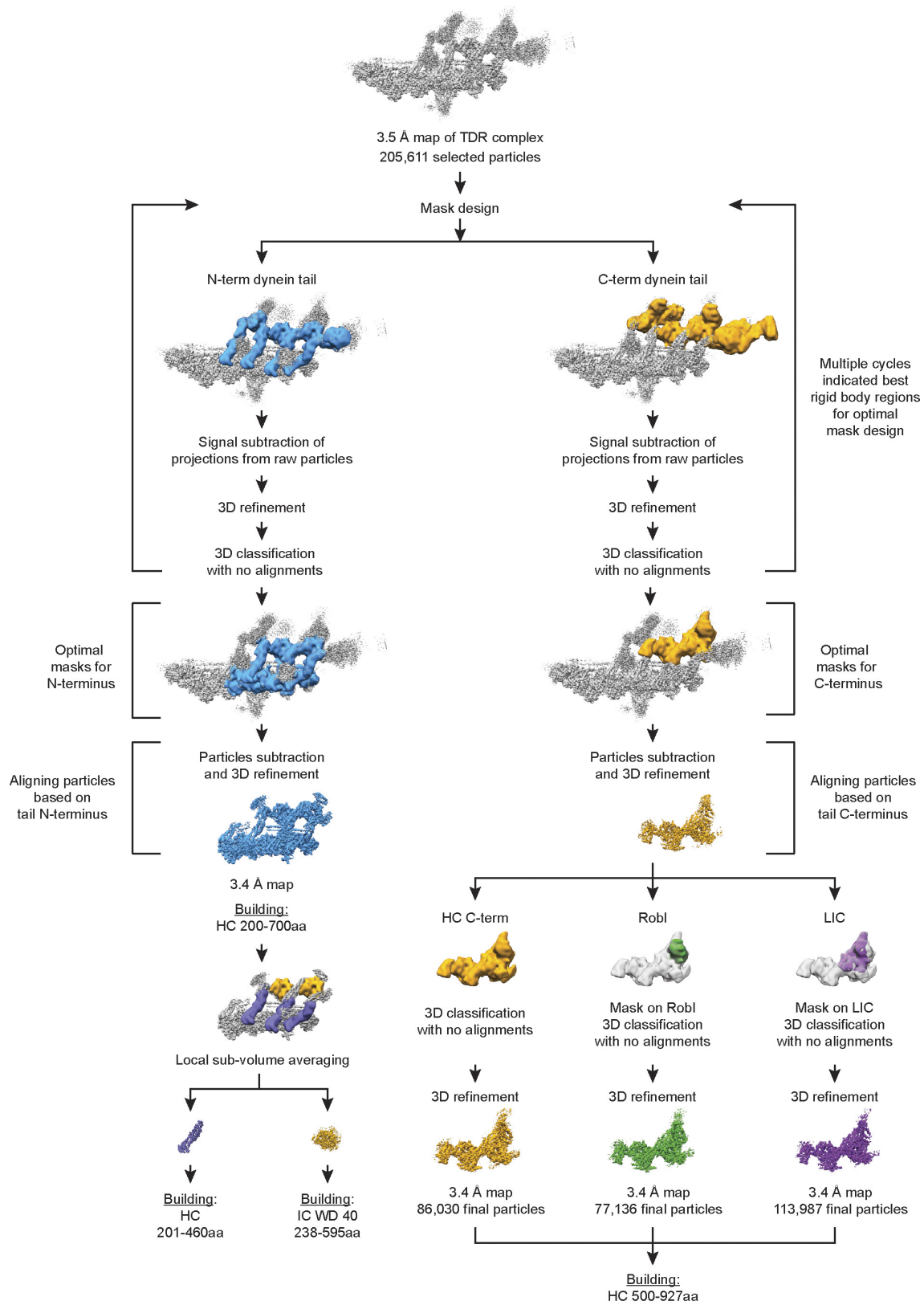
**a**, A one-cumulative frequency distribution plot showing run-lengths of DDB, DDR and DDH, with fit to a one-phase exponential decay. The decay constant (run length) and  $R^2$  value (least squares regression) of the fit are shown. We measured 785, 677 and 684 events for DDB, DDH and DDR, respectively, from microtubules of at least  $20 \mu\text{m}$  in length from three chambers. **b–f**, Distribution of mean velocities of processive (unidirectional, minus-end-directed) events for DDB ( $n = 3,343$ ) and DDR ( $n = 3,162$ ) (**b**); DDB and DDH ( $n = 3,744$ ) (**c**); active mutant dynein

in complex with dynactin and BICD2 (mtDDB,  $n = 905$ ) or BICDR1 (mtDDR,  $n = 1,183$ ) (**d**); the colocalized mtDDR complexes containing both TMR-dynein tail and Alexa Fluor 647-full-length dynein (tail-dynein,  $n = 939$ ) or Alexa Fluor 647-only complexes containing only full-length dynein (dynein-only,  $n = 1,004$ ) (**e**); and all DDB complexes and complexes with both fluorophores, and hence two dyneins (colocalizers,  $n = 660$ ) (**f**). Mean  $\pm$  s.e.m. values were estimated by fitting the histograms to a Gaussian distribution (dashed lines).

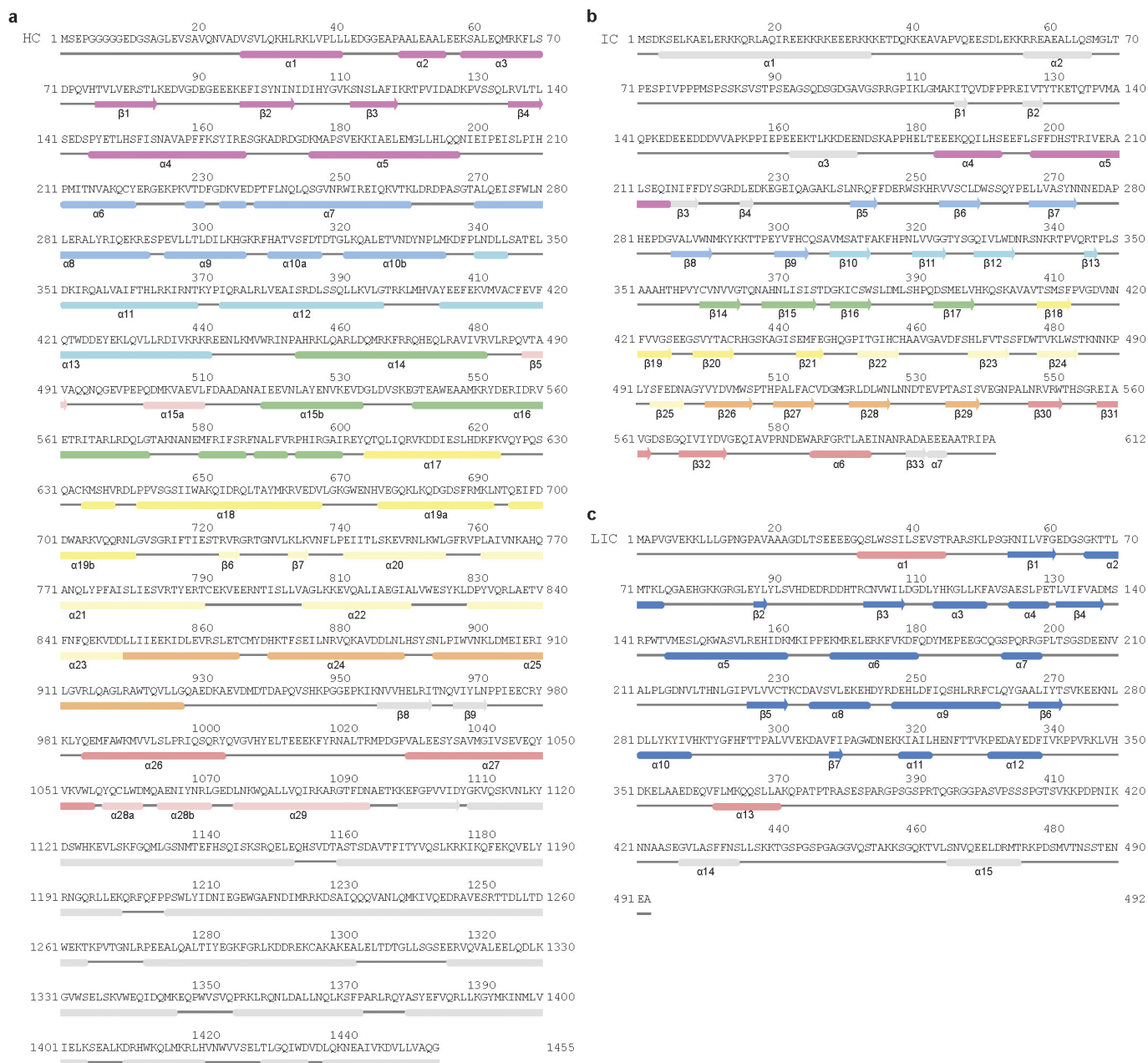


**Extended Data Figure 3 | Single-particle cryo-EM analysis of TDR complex at 3.5 Å resolution.** **a**, Micrograph of the TDR complex (representative of 26,906 micrographs). **b**, Typical 2D-class averages of TDR in different orientations. **c**, The overall density map of TDR was analysed by ResMap, showing a resolution distribution from

3 to 8 Å. **d**, The gold-standard FSC curve of the overall TDR map. **e**, Mesh representation of 3.4 Å resolution density map of  $\alpha$  helices from dynein-B1 obtained by focused 3D classification and refinement. **f**, Sample density obtained by local sub-volume averaging, showing  $\beta$  strands from IC WD40.



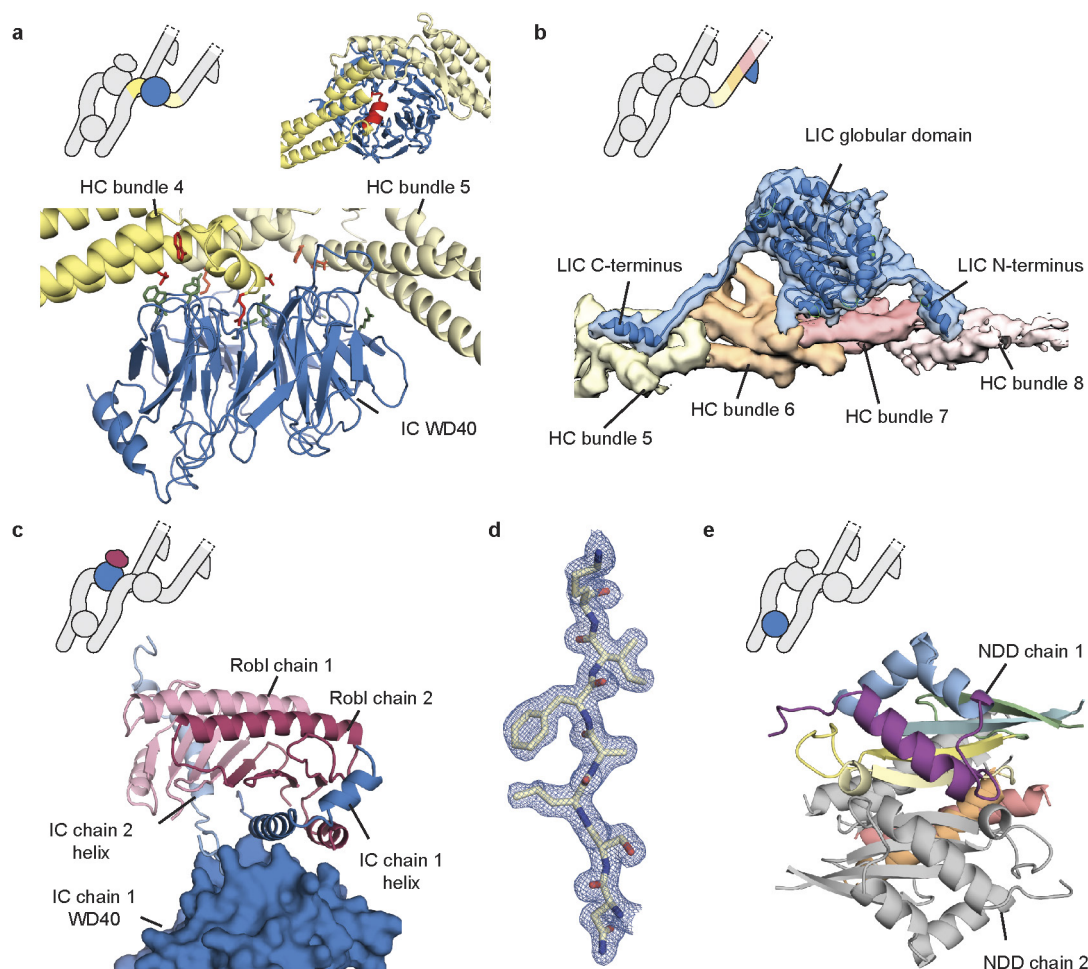
**Extended Data Figure 4 | Cryo-EM data procedures of TDR.** Focused 3D classification and refinement procedures used in this study to improve density maps for dynein tails.



### Extended Data Figure 5 | Secondary structure diagram of dynein HC.

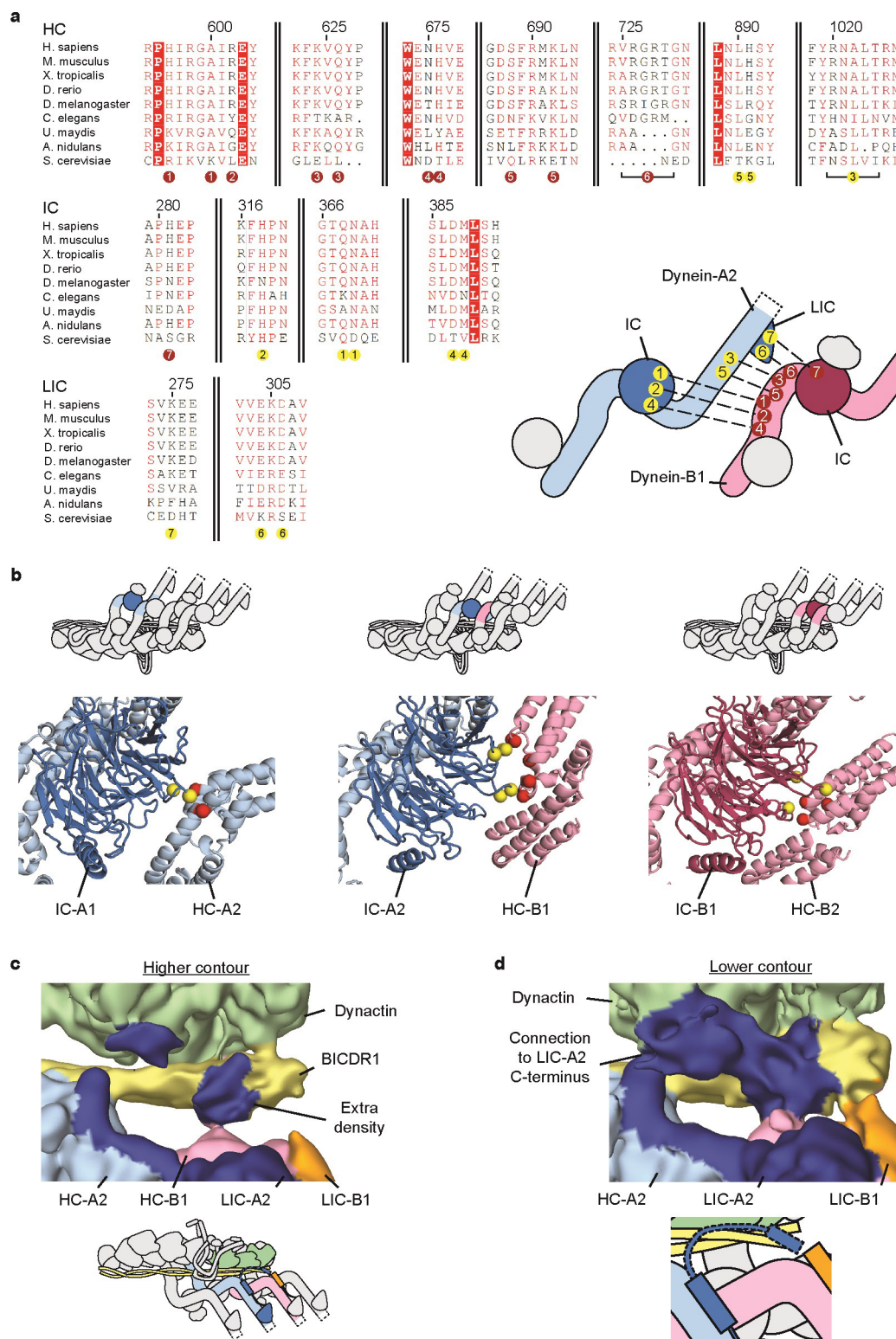
**a**, Secondary structure elements of dynein HC are matched against the primary sequence showing the NDD (purple) and the dynein helical bundles (blue; cyan; green; yellow; pale yellow; orange; red; pink). **b**, Secondary structure elements of IC. Extended N-terminal regions are coloured purple and other elements are coloured according to the blade of

the WD40 domain to which they belong, except sheet  $\beta 5$ , which associates with  $\beta 30$ – $\beta 32$ . **c**, Secondary structure elements of LIC, showing the globular domain helices and sheets (blue) and the two helices that pack against the HC (red). Jpred<sup>53</sup> secondary structure predictions of features not seen in the electron microscopy map are shown in grey.



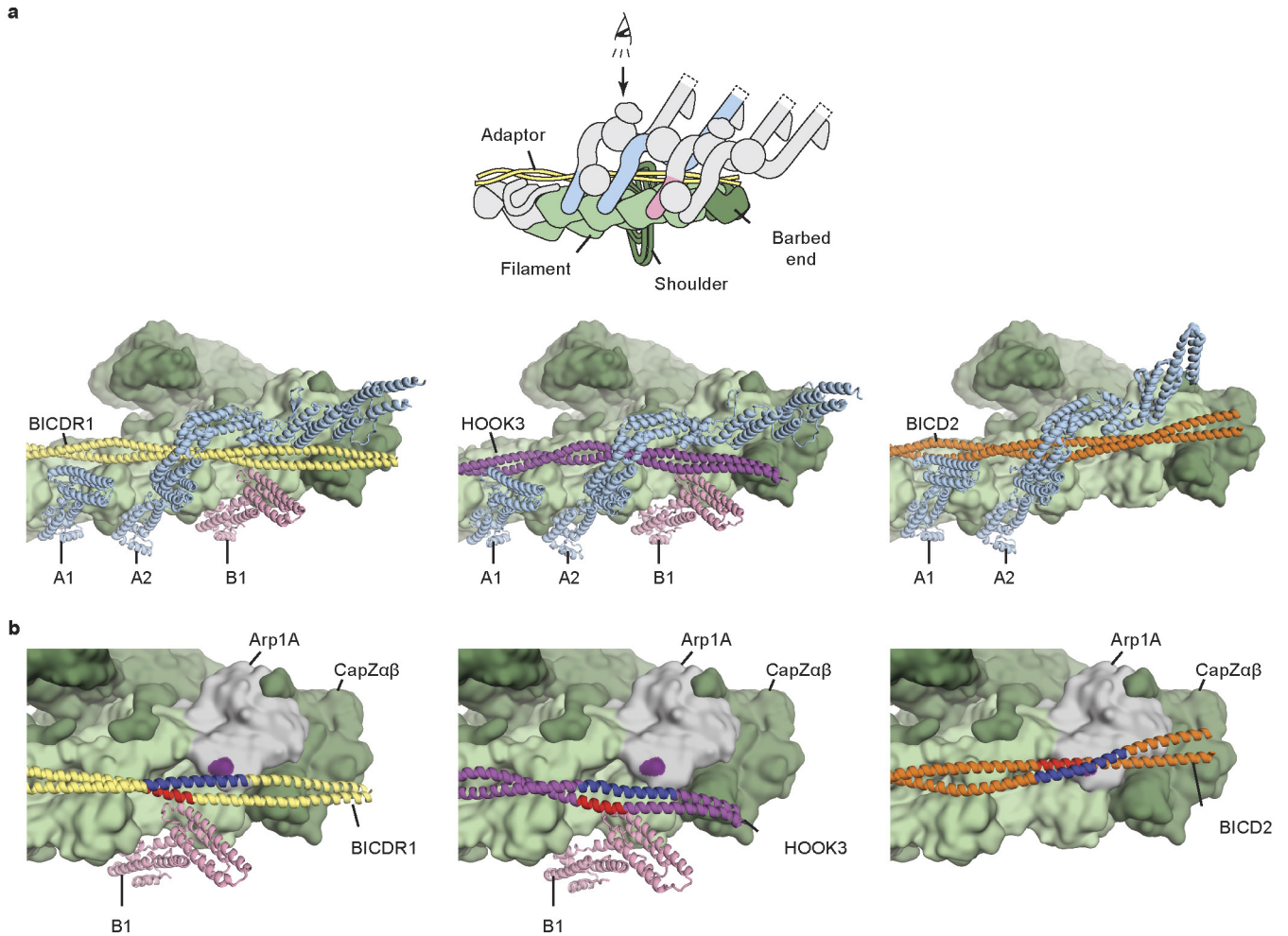
**Extended Data Figure 6 | Interactions between dynein subunits. a,** The dynein HC (yellow) interacts with the IC WD40 domain (blue) using bundles 4 and 5, with a helical segment (red cartoon) sitting in the WD40 central cavity. Dynein-A2 is shown. Interacting residues are shown as sticks (bottom panel), with HC residues in red and IC residues in green. **b,** Density map and model showing how the LIC (density and cartoon, blue) N- and C-terminal regions extend from the globular domain and

pack against the HC (density, coloured by bundle number). Dynein-A2 is shown. **c,** ROBL1 (cartoon, light and dark pink) makes contacts with the IC N-terminal helices (cartoon, light and dark blue), which mediate the interaction between ROBL1 and the IC WD40 (surface). **d,** Representative density from the 1.9 Å resolution NDD crystal structure. **e,** Cartoon model of the NDD showing one chain in rainbow spectrum.



**Extended Data Figure 7 | Dynein–dynein contacts and interactions at the BICDR1 N terminus.** **a**, Conservation diagram showing sequence similarity between A2 and B1 interacting residues. Residues coloured white with red background are completely conserved, whereas residues coloured red show sequence similarity at that position. Residues at each interaction site are numbered below the alignment (A2 residues in yellow circles, B1 residues in red circles). These numbers label the accompanying cartoon to show the dynein chains that constitute each interaction. Alignment generated by ESPrpt<sup>54</sup> (<http://esprpt.icbp.fr>). **b**, Intermediate

chain interactions showing connections between the IC of A1 and the HC of A2; the IC of A2 and the HC of B1; and the IC of B1 and the HC of B2. Interacting sites on each IC are shown as yellow spheres; sites on each HC are shown as red spheres. **c**, B1 (pink) contacts extra density (labelled, blue) adjacent to the BICDR1 coiled coil. The cartoon below shows the location of the area depicted (correspondingly coloured). **d**, Weak density connects the extra density with the LIC A2 helix 13 (blue). A cartoon representation of the area depicted is shown below.



**Extended Data Figure 8 | Comparison between different adaptors recruiting dynein. a,** The TDR structure (left) is compared to models of TDH (middle) and TDB (right). Although the paths of BICDR1 (yellow), HOOK3 (magenta) and BICD2 (orange) vary along the surface of dynactin (green surface), dynein-A HCs (light blue) bind at the same sites in each

complex. **b,** Zoomed-in views of the barbed end of dynactin show that BICD2 adopts an upwards position to contact ARP1A (grey), whereas BICDR1 and HOOK3 adopt lower positions to bind dynein-B using the region coloured in red. The BICD2–ARP1A interaction site is highlighted in purple.

**Extended Data Table 1 | Cryo-EM data collection parameters of TDR and TDH structures and model refinement statistics of the 3.5 Å resolution TDR structure**

|  | TDR_1  | TDR_2   | TDH    |
|--|--------|---------|--------|
| <b>Data collection and processing</b>  |        |         |        |
| Voltage (kV)                           | 300    | 300     | 300    |
| Electron exposure (e-/Å <sup>2</sup> ) | 52     | 52      | 45     |
| Pixel size (Å)                         | 1.34   | 1.34    | 1.42   |
| Number of sessions                     | 1      | 11      | 5      |
| Micrographs                            | 2,459  | 26,906  | 5,464  |
| Symmetry imposed                       | C1     | C1      | C1     |
| Final particle images (no.)            | 12,420 | 205,611 | 23,407 |
| Map resolution (Å)                     | 6.5    | 3.5     | 6.7    |
| FSC threshold                          | 0.143  | 0.143   | 0.143  |

|  | N-terminal tail | C-terminal tail | TDR      | TDR(ordered) |
|--|-----------------|-----------------|----------|--------------|
| <b>Refinement</b>                                |                 |                 |          |              |
| Map  | EMD-4169        | EMD-4170        | EMD-4168 | EMD-4168     |
| Map resolution (Å)                               | 3.4             | 3.4             | 8        | 3.5          |
| FSC threshold                                    | 0.143           | 0.143           | 0.143    | 0.143        |
| Map sharpening <i>B</i> factor (Å <sup>2</sup> ) | -50             | -30             | 0        | -70          |
| Map CC (around atoms)                            | 0.76            | 0.70            | 0.79     | 0.65         |
| <b>Model composition</b>                         |                 |                 |          |              |
| Non-hydrogen atoms                               | 28,871          | 5,164           | 92,789   | 90,725       |
| Protein residues                                 | 3,555           | 628             | 13,982   | 13,567       |
| Ligands (ADP/ATP)                                | 0/0             | 0/0             | 9/1      | 9/1          |
| <b>R.m.s. deviations</b>                         |                 |                 |          |              |
| Bond lengths (Å)                                 | 0.02            | 0.01            | 0.01     | 0.01         |
| Bond angles (°)                                  | 1.59            | 1.93            | 1.60     | 1.60         |
| <b>Validation</b>                                |                 |                 |          |              |
| MolProbity score                                 | 2.12            | 2.21            | 2.05     | 2.05         |
| Clashscore                                       | 11.18           | 13.19           | 9.08     | 9.08         |
| Poor rotamers (%)                                | 0.29            | 0.36            | 0.44     | 0.44         |
| <b>Ramachandran plot</b>                         |                 |                 |          |              |
| Favored (%)                                      | 89.83           | 88.75           | 89.13    | 89.17        |
| Disallowed (%)                                   | 0.11            | 0.16            | 0.14     | 0.14         |
| C $\beta$ deviations (%)                         | 0.00            | 0.00            | 0.00     | 0.00         |

The TDR\_1 dataset is included in TDR\_2 dataset. The N-terminal tail model consists of HC of A2 (residues 201–829), HC of B1 (residues 201–629), HC of B2 (residues 201–575), IC of A2 WD40 domain, BICDR1 (residues 132–210), ARP1B, ARP1D, ARP1F, CAPZ $\alpha$  and CAPZ $\beta$ . The C-terminal tail model consists of HC of A2 (residues 517–927) and HC of B1 (residues 453–702). The TDR (ordered) model consists of all parts of TDR for which density was seen.

**Extended Data Table 2 | Crystal structure data collection parameters and model refinement statistics of the 1.9 Å resolution structure of the human dynein NDD**

|  | NDD (PDB 5OWO)                                 |
|--|--|
| <b>Data collection</b>                               |  |
| Space group  | P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell dimensions                                      |  |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                   | 50.5, 101.8, 176.19                            |
| $\alpha$ , $\beta$ , $\gamma$ (°)                    | 90, 90, 90                                     |
| Resolution (Å)                                       | 50.45-1.86 (1.94-1.86) *                       |
| <i>R</i> <sub>merge</sub>                            | 0.291(3.939)                                   |
| <i>I</i> / $\sigma$ <i>I</i>                         | 3.9(0.4)                                       |
| Completeness (%)                                     | 94.3(98.5)                                     |
| Redundancy   | 3.1 (2.9)                                      |
| <b>Refinement</b>                                    |  |
| Resolution (Å)                                       | 1.80   |
| No. reflections                                      | 48901  |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>  | 27.09/29.24                                    |
| <i>B</i> -factor, from Wilson plot (Å <sup>2</sup> ) | 25.7   |
| R.m.s. deviations                                    |  |
| Bond lengths (Å)                                     | 0.02   |
| Bond angles (°)                                      | 2.15   |

\*Values in parentheses are for the highest-resolution shell.

# A mildly relativistic wide-angle outflow in the neutron-star merger event GW170817

K. P. Mooley<sup>1,2,3</sup>, E. Nakar<sup>4</sup>, K. Hotokezaka<sup>5</sup>, G. Hallinan<sup>3</sup>, A. Corsi<sup>6</sup>, D. A. Frail<sup>2</sup>, A. Horesh<sup>7</sup>, T. Murphy<sup>8,9</sup>, E. Lenc<sup>8,9</sup>, D. L. Kaplan<sup>10</sup>, K. De<sup>3</sup>, D. Dobie<sup>8,9,11</sup>, P. Chandra<sup>12,13</sup>, A. Deller<sup>14</sup>, O. Gottlieb<sup>4</sup>, M. M. Kasliwal<sup>3</sup>, S. R. Kulkarni<sup>3</sup>, S. T. Myers<sup>2</sup>, S. Nissanke<sup>15</sup>, T. Piran<sup>7</sup>, C. Lynch<sup>8,9</sup>, V. Bhallerao<sup>16</sup>, S. Bourke<sup>17</sup>, K. W. Bannister<sup>11</sup> & L. P. Singer<sup>18</sup>

GW170817 was the first gravitational wave detection of a binary neutron-star merger<sup>1</sup>. It was accompanied by radiation across the electromagnetic spectrum and localized<sup>2</sup> to the galaxy NGC 4993 at a distance of 40 megaparsecs. It has been proposed that the observed  $\gamma$ -ray, X-ray and radio emission is due to an ultra-relativistic jet launched during the merger, directed away from our line of sight<sup>3–6</sup>. The presence of such a jet is predicted from models that posit neutron-star mergers as the central engines that drive short hard  $\gamma$ -ray bursts<sup>7,8</sup>. Here we report that the radio light curve of GW170817 has no direct signature of an off-axis jet afterglow. Although we cannot rule out the existence of a jet pointing elsewhere, the observed  $\gamma$ -rays could not have originated from such a jet. Instead, the radio data require a mildly relativistic wide-angle outflow moving towards us. This outflow could be the high-velocity tail of the neutron-rich material dynamically ejected during the merger or a cocoon of material that breaks out when a jet transfers its energy to the dynamical ejecta. The cocoon model explains the radio light curve of GW170817 as well as the  $\gamma$ -rays and X-rays (possibly also ultraviolet and optical emission)<sup>9–15</sup>, and is therefore the model most consistent with the observational data. Cocoons may be a ubiquitous phenomenon produced in neutron-star mergers, giving rise to a heretofore unidentified population of radio, ultraviolet, X-ray and  $\gamma$ -ray transients in the local Universe.

The radio discovery<sup>12</sup> of GW170817, as well as observations within the first month post-merger, were interpreted in the framework of classical off-axis jet, cocoon, and dynamical ejecta. We continued to observe GW170817 with the Karl G. Jansky Very Large Array (VLA), the Australia Telescope Compact Array (ATCA) and the upgraded Giant Metrewave Radio Telescope (uGMRT), spanning the frequency range 0.6–18 GHz, whilst optical and X-ray telescopes were constrained by proximity to the Sun. Our radio detections span up to 107 days post-merger (Figure 1 and Methods). These data show a steady rise in the radio light curve and a spectrum consistent with optically-thin synchrotron emission. A joint temporal and spectral power-law fit to these data of the form  $S \propto \nu^\alpha t^\delta$  is well-described by a spectral index  $\alpha = -0.6$  and a temporal index  $\delta = +0.8$  (see Methods). On 2017 November 18 (93 days post-merger) the peak luminosity at 1.6 GHz was  $2 \times 10^{27}$  erg s<sup>-1</sup> Hz<sup>-1</sup>, a luminosity undetectable for even the nearest short-hard  $\gamma$ -ray burst (SGRB) afterglow discovered to date<sup>16</sup>.

The (sub-luminous) gamma-ray emission detected immediately after the gravitational wave detection<sup>17</sup> must have been emitted by a

relativistic outflow<sup>14</sup>, but an on-axis jet (scenario A in Figure 2) was ruled out by the late turn-on of the X-ray and radio emission<sup>3–6,11–13</sup>. If GW170817 produced a regular (luminous) SGRB pointing away from us, then the interaction of the jet with the circum-merger medium would have decelerated the jet, and the afterglow emission would have eventually entered into our line of sight, thus producing a so-called off-axis afterglow<sup>18,19</sup>. For this geometry, the light curve rises sharply and peaks when the jet Lorentz factor  $\gamma \sim 1/(\theta_{\text{obs}} - \theta_j)$ , and then undergoes a power law decline ( $\theta_{\text{obs}}$  is the angle between the jet axis and the line of sight, and  $\theta_j$  is the jet opening angle). This behavior is clearly inconsistent with the full light curve shown in Figure 1. The rise is less steep than an off-axis jet and it is consistent with a monotonic increase without either a plateau or a subsequent decay. Initial off-axis models (based on available X-ray and radio data at the time) predicted a radio flux density<sup>3–5,12</sup> of  $\sim 10 \mu\text{Jy}$  (between 3 GHz and 10 GHz)  $\sim 100$  days post-merger, while our measured values are at least a factor of five larger. The discrepancy with the off-axis jet model is further demonstrated in Figure 3 where various jet and medium parameters are considered, showing in all cases a similar general light curve shape which cannot fit the data. We have considered a wide range in the phase space of off-axis models, and can rule out an off-axis jet (scenario B in Figure 2) as the origin of the radio afterglow of GW170817. We show below that even if we consider a “structured jet”, in which the outflow has an angular dependence of the Lorentz factor and energy (scenario E in Figure 2 represents one such configuration), the observed radiation arises predominantly from a mildly relativistic outflow moving towards us (at an angle less than  $1/\gamma$ ), and we do not detect the observational signature of a relativistic core within the structured jet.

With a highly collimated off-axis jet ruled out, we next consider spherical or quasi-spherical ejecta components. A single spherical shell of expanding ejecta will produce a light curve that rises as  $S \sim t^3$ . The light curve of GW170817 immediately rules out such a simple single-velocity ejecta model. The gradual but monotonic rise seen in our radio data ( $S \propto t^{0.8}$ ; Figure 1) points instead to *on-axis* emission from a mildly relativistic blast wave where the energy is increasing with time (due to more mass residing in slower ejecta, which is seen at later times). For example, using canonical microphysical parameters ( $\epsilon_B = 0.01$ ,  $\epsilon_e = 0.1$ ), a density of  $10^{-4} \text{ cm}^{-3}$  implies that between day 16 to day 107 the blast wave decelerates from  $\gamma \sim 3.5$  to  $\gamma \sim 2.5$  and its isotropic equivalent energy increases from  $\sim 10^{49}$  erg to  $\sim 10^{50}$  erg.

<sup>1</sup>Hintze Fellow, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. <sup>2</sup>National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA. <sup>3</sup>California Institute of Technology, 1200 East California Boulevard, MC 249-17, Pasadena, California 91125, USA. <sup>4</sup>The Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. <sup>5</sup>Department Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA. <sup>6</sup>Department of Physics and Astronomy, Texas Tech University, Box 41051, Lubbock, Texas 79409-1051, USA. <sup>7</sup>Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. <sup>8</sup>Sydney Institute for Astronomy, School of Physics, University of Sydney, New South Wales 2006, Australia. <sup>9</sup>ARC Centre of Excellence for All-sky Astrophysics (CAASTRO). <sup>10</sup>Department of Physics, University of Wisconsin - Milwaukee, Milwaukee, Wisconsin 53201, USA. <sup>11</sup>ATNF, CSIRO Astronomy and Space Science, PO Box 76, Epping, New South Wales 1710, Australia. <sup>12</sup>National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Pune University Campus, Ganeshkhind Pune 411007, India. <sup>13</sup>Department of Astronomy, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden. <sup>14</sup>Centre for Astrophysics and Supercomputing, Swinburne University of Technology, John Street, Hawthorn, Victoria 3122, Australia. <sup>15</sup>Institute of Mathematics, Astrophysics and Particle Physics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands. <sup>16</sup>Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India. <sup>17</sup>Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, S-439 92 Onsala, Sweden. <sup>18</sup>Astroparticle Physics Laboratory, NASA Goddard Space Flight Center, Mail Code 661, Greenbelt, Maryland 20771, USA.

On the other hand, a density of  $0.01 \text{ cm}^{-3}$  implies a velocity range of  $0.8c$  to  $0.65c$  and energy that rises from  $10^{48} \text{ erg}$  to  $10^{49} \text{ erg}$ . Figure 4 shows that a quasi-spherical outflow with a velocity profile  $E(>\beta\gamma) \propto (\beta\gamma)^{-5}$  provides an excellent fit to the data (see Methods), and it is almost independent of the assumed circum-merger density and micro-physical parameters. The energy injection into the blast wave during the time span of the observations (day 16 to day 107) increases its energy by a factor of  $\sim 10$ . The possible origin of the outflow depends on its energy and velocity. A faster and more energetic outflow, with  $\gamma \sim 2\text{--}3$  and energy of  $10^{49}\text{--}10^{50} \text{ erg}$ , is a natural outcome of the cocoon driven by a wide-angle choked jet<sup>9,11,14</sup> (scenario C in Figure 2). This scenario explains many of the puzzling characteristics of GW170817. First, the breakout of the cocoon from the ejecta can produce the observed sub-luminous gamma-ray signal, including its peak energy and spectral evolution<sup>14</sup> (see also Methods). Second, it provides a natural explanation for the high velocities of the bulk of the ejecta ( $\sim 0.3c$ ) and for the early bright UV and optical light<sup>11,13,15</sup>. On the other hand, a slower and less energetic outflow, with  $\beta \approx 0.8\text{--}0.6$  ( $\gamma \sim 1.67\text{--}1.25$ ) and energy of  $10^{48}\text{--}10^{49} \text{ erg}$  can arise from the fast tail of the merger ejecta<sup>20–22</sup> (scenario D in Figure 2), although we note that this component cannot explain the gamma-ray signal (GRB 170817A) from GW170817. These two scenarios can be easily distinguished by Very Long Baseline Interferometry or monitoring of the radio evolution on  $\sim$  years timescale.

A hidden jet, which does not contribute significantly to the observed afterglow, may still exist (scenario E in Figure 2), but its properties are tightly constrained. First, its edge must be far enough from the line-of-sight ( $\gtrsim 10$  degrees), which rules out off-axis gamma-ray emission as the source of GRB 170817A. Second, for every reasonable set of parameters, an off-axis jet would have been brighter than the fast tail of the ejecta, implying that the observed emission must be dominated by a  $\gamma \sim 2\text{--}3$  outflow (i.e. a cocoon) for the jet to remain undetected. In addition, the jet energy should, most likely, be much lower than that of the cocoon, which needs fine-tuning of the jet properties (see Methods). We therefore conclude from the lack of a signature from an off-axis jet, that the jet was likely choked (scenario C in Figure 2).

We compared the 3 GHz radio and X-ray<sup>4–6</sup> detections obtained on 2017 September 02–03 (15–16 days post-merger). The measurements at these two disparate frequencies imply a spectral index of  $-0.6$ , consistent with our multi-epoch, multi-frequency, radio-only measurements (see Methods and Extended Data Figure 4). It is therefore likely that the radio and X-rays originate from the same (synchrotron) source, viz. a mildly relativistic outflow. This common origin can be confirmed if the X-ray flux continues to rise in a similar manner as the radio. We also highlight that, while at early times the cooling break will lie well above the soft X-ray frequencies, beyond  $\sim 10^2\text{--}10^3$  days post-merger this break may be seen moving downwards in frequency within the electromagnetic spectrum. If the cooling break stays above  $10^{18} \text{ Hz}$ , the common origin of the radio and X-rays implies that the Chandra telescope will detect a brighter X-ray source (flux between  $0.7 \times 10^{-14}$  and  $5.2 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$  in the  $0.3\text{--}10 \text{ keV}$  band; see Methods) during its observation of GW170817 on December 03–06 (note: subsequent to the submission of this paper, the X-ray observations took place and confirmed our prediction). If a different spectral index is derived from these X-ray observations relative to the in-band radio spectral index presented here, or indeed at any time within  $\sim 1000$  days of the merger, it will indicate that the cooling break has already shifted below the X-ray band, which would favor the fast tail of the merger ejecta as the common source of the X-ray and radio emission (see Methods).

The confirmation of a wide-angle outflow in GW170817 bodes well for electromagnetic counterpart searches of future gravitational wave events. Although on-axis (and slightly off-axis;  $\theta_{\text{obs}} < 20$  degrees) jets produce bright panchromatic afterglows, they represent only a small fraction ( $\sim 10\%$ ) of the gravitational wave events (factoring in the larger detectable distance for face-on events<sup>23</sup>). In contrast, the emission from wide-angle cocoons<sup>9–11</sup> will be potentially seen in a much

larger fraction of events, and at virtually all wavelengths, thus increasing the probability of the detection of electromagnetic counterparts. The radio emission from the cocoon, evolving on timescales of weeks to months, especially provides a distinct signature (as opposed to the more common supernovae and AGN transients) and diagnostics for observers. Specifically in the case of GW170817, continued monitoring of the radio light curve will provide an independent constraint on the circum-merger density and thereby the properties of the blast wave that dominated the early-time radio emission.

Our radio data support the hypothesis of a choked jet giving rise to a mildly relativistic cocoon (scenario C in Figure 2), but this is only one of the possible outcomes of neutron star merger events (see Figure 2). In some cases, the jet may break out after depositing a fraction of its energy into the cocoon, thereby still successfully producing a SGRB<sup>11</sup> (scenario E in Figure 2). Indeed, a plateau in the distribution of SGRB durations has been highlighted as evidence that SGRB jets often propagate through slower traveling ejecta before breakout and at times it is choked<sup>24</sup>. The relative fractions of neutron star mergers that successfully produce a SGRB or a choked jet can be directly probed via radio follow-up of a sample of neutron star mergers in the upcoming LIGO-Virgo campaigns.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 28 November; accepted 11 December 2017.**

**Published online 20 December 2017.**

- Abbott, B. P., Abbott, R., Abbott, T. D., *et al.* GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.*, **119**, 161101 (2017)
- Abbott, B. P., Abbott, R., Abbott, T. D., *et al.* Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J. Lett.*, **848**, L13 (2017)
- Alexander, K. D., Berger, E., Fong, W., *et al.* The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. VI. Radio Constraints on a Relativistic Jet and Predictions for Late-time Emission from the Kilonova Ejecta. *Astrophys. J. Lett.*, **848**, L21 (2017)
- Margutti, R., Berger, E., Fong, W., *et al.* The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. V. Rising X-Ray Emission from an Off-axis Jet. *Astrophys. J. Lett.*, **848**, L20 (2017)
- Troja, E., Piro, L., van Eerten, H., *et al.* The X-ray counterpart to the gravitational-wave event GW170817. *Nature*, **551**, 71–74
- Haggard, D., Nynka, M., Ruan, J. J., Kalogera, V., Cenko, S. B., Evans, P. & Kennea, J. A. A Deep Chandra X-Ray Study of Neutron Star Coalescence GW170817. *Astrophys. J. Lett.*, **848**, L25 (2017)
- Eichler, D., Livio, M., Piran, T. & Schramm, D. N., Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars, *Nature*, **340**, 126–128 (1989)
- Nakar, E., Short-hard gamma-ray bursts. *Phys. Rep.*, **442**, 166–236 (2007)
- Gottlieb, O., Nakar, E. & Piran, T., The cocoon emission - an electromagnetic counterpart to gravitational waves from neutron star mergers. *Mon. Not. R. Astron. Soc.*, **473**, 576–584 (2018)
- Lazzati, D., Deich, A., Morsony, B. J. & Workman, J. C. Off-axis emission of short  $\gamma$ -ray bursts and the detectability of electromagnetic counterparts of gravitational-wave-detected binary mergers. *Mon. Not. R. Astron. Soc.*, **471**, 1652–1661 (2017)
- Kasliwal, M. M., Nakar, E., Singer, L. P., *et al.* Illuminating Gravitational Waves: A Concordant Picture of Photons from a Neutron Star Merger, Science DOI: 10.1126/science.aap9455 (2017)
- Hallinan, G., Corsi, A., Mooley, K. P., *et al.* A Radio Counterpart to a Neutron Star Merger, Science DOI: 10.1126/science.aap9855 (2017)
- Evans, P. A., Cenko, S. B., Kennea, J. A. *et al.*, Swift and NuSTAR observations of GW170817: Detection of a blue kilonova, Science DOI: 10.1126/science.aap9580 (2017)
- Gottlieb, O., Nakar, E., Piran, T. & Hotokezaka, K., A cocoon shock breakout as the origin of the gamma-ray emission in GW170817. ArXiv:171005896G (2017)
- Piro, A. L. & Kollmeier, J. A. Evidence for Cocoon Emission from the Early Light Curve of SSS17a. ArXiv:1710.05822 (2017)
- Fong, W., Berger, E., Margutti, R. & Zauderer, B. A., A Decade of Short-duration Gamma-Ray Burst Broadband Afterglows: Energetics, Circumburst Densities, and Jet Opening Angles. *Astrophys. J.*, **815**, 102 (2015)
- Abbott, B. P., Abbott, R., Abbott, T. D., *et al.*, Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *Astrophys. J. Lett.*, **848**, L13 (2017)
- Granot, J., Panaitescu, A., Kumar, P. & Woosley, S. E. Off-Axis Afterglow Emission from Jetted Gamma-Ray Bursts. *Astrophys. J. Lett.*, **570**, L61–L64 (2002)

19. Nakar, E., Piran, T. & Granot, J., The Detectability of Orphan Afterglows, *Astrophys. J.*, **579**, 699–705 (2002)
20. Bauswein, A., Goriely, S. & Janka, H.-T. Systematics of Dynamical Mass Ejection, Nucleosynthesis, and Radioactively Powered Electromagnetic Signals from Neutron-star Mergers. *Astrophys. J.*, **773**, 78 (2013)
21. Hotokezaka, K., Kiuchi, K., Kyutoku, K., *et al.* Mass ejection from the merger of binary neutron stars. *Phys. Rev. D*, **87**, 4001 (2013)
22. Kyutoku, K., Ioka, K. & Shibata, M., Ultrarelativistic electromagnetic counterpart to binary neutron star mergers. *Mon. Not. R. Astron. Soc. Lett.*, **437**, L6–L10 (2014)
23. Nissanke, S., Kasliwal, M. & Georgieva, A., Identifying Elusive Electromagnetic Counterparts to Gravitational Wave Mergers: An End-to-end Simulation. *Astrophys. J.*, **767**, 124 (2013)
24. Moharana, R. & Piran, T. Observational evidence for mass ejection accompanying short gamma-ray bursts. *Mon. Not. R. Astron. Soc.*, **472**, L55–L59 (2017)

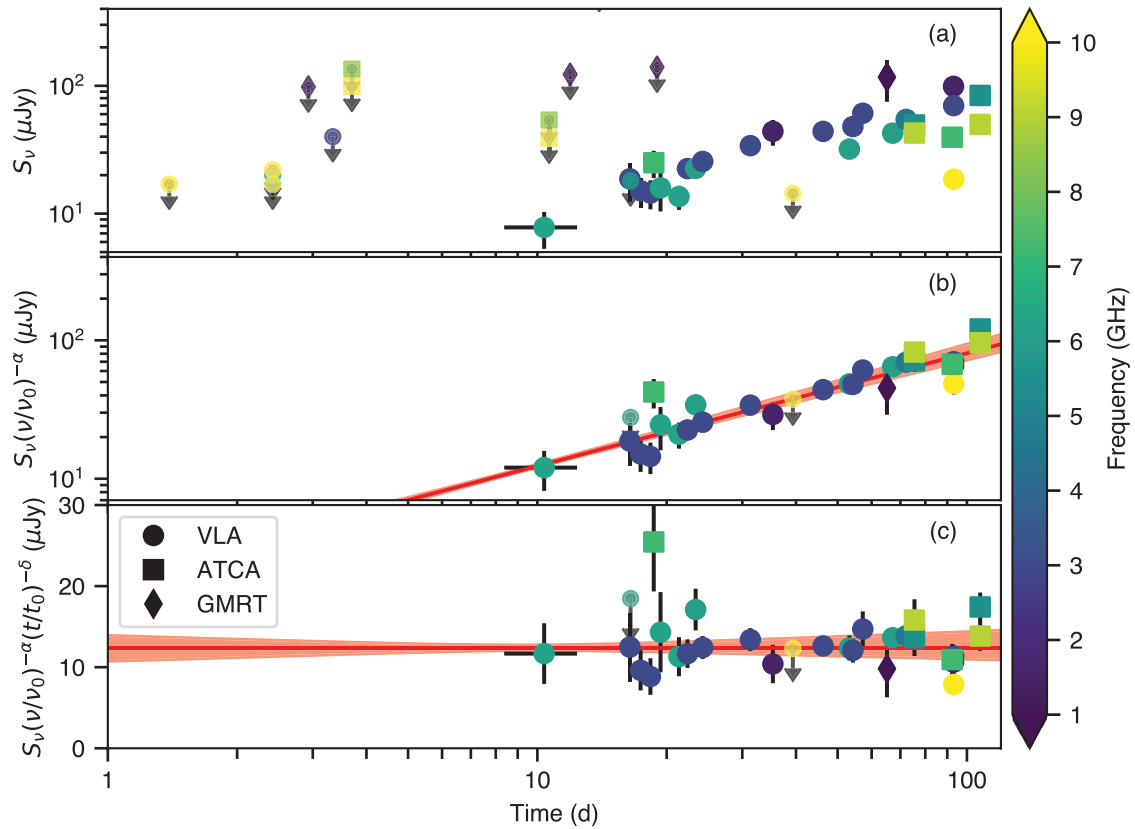
**Acknowledgements** We acknowledge the support and dedication of the staff of the National Radio Astronomy Observatory and particularly thank the VLA Director, M. McKinnon, as well as A. Mioduszewski and H. Medlin, for making the VLA campaign possible. We thank Britt Griswold (NASA/GSFC) for beautiful graphic arts. SK thanks M. Shull for discussions. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. We thank the GMRT staff for scheduling our observations. The GMRT is run by the National Centre for Radio Astrophysics of the Tata Institute of Fundamental Research. The Australia Telescope Compact Array is part of the Australia Telescope National Facility which is funded by the Australian Government for operation as a National Facility managed by CSIRO. KM's research is supported by the Hintze Centre for Astrophysical Surveys which is funded through the Hintze Family Charitable Foundation. EN acknowledges the support of an ERC starting grant (GRB/SN) and an ISF grant (1277/13). GH acknowledges the support of NSF award AST-1654815. AC acknowledges support from the National Science Foundation CAREER award #1455090 titled 'CAREER: Radio and gravitational-wave emission from the largest explosions since the Big Bang'. AH acknowledges support by the I-Core Program of the Planning and Budgeting Committee and the Israel Science Foundation. TM acknowledges the support of the Australian Research Council through grant FT150100099. Parts of this research were conducted by the Australian Research Council Centre of Excellence for

All-sky Astrophysics (CAASTRO), through project number CE110001020. DK was supported by NSF grant AST-1412421. MK's work was supported by the GROWTH (Global Relay of Observatories Watching Transients Happen) project funded by the National Science Foundation under PIRE Grant No 1545949. This work is part of the research program Innovational Research Incentives Scheme (Vernieuwingsimpuls), which is financed by the Netherlands Organization for Scientific Research through the NWO VIDI Grant No. 639.042.612-Nissanke and NWO TOP Grant No. 62002444--Nissanke. PC acknowledges support from the Department of Science and Technology via SwarnaJayanti Fellowship awards (DST/SJF/PSA-01/2014-15). TP acknowledges the support of Advanced ERC grant TRex. VB acknowledges the support of the Science and Engineering Research Board, Department of Science and Technology, India, for the GROWTH-India project.

**Author Contributions** KM, EN, KH, GH and DF wrote the paper. AC compiled the references. AC and AH compiled the methods section. DD and KD compiled the radio measurements table. KM managed the VLA observing program and processed all the VLA data. SM, AD and SB helped plan the VLA observations. EN, KH, DK and KM prepared the figures. TM planned and managed ATCA observations and data analysis and contributed to the manuscript text. DK helped propose for and plan the ATCA observations and contributed to the manuscript text. EL, DD, CL and KB helped with ATCA observations and data reduction. KD planned and managed GMRT observations and contributed to manuscript text. KM and PC processed the GMRT data. VB helped in the GMRT observations. OG and EN provided the cocoon simulation. KH provided the spherical ejecta model. SN did the GW and cocoon rates analysis. SK, TP, MK and LS provided text for the paper. All coauthors discussed the results and provided comments on the manuscript.

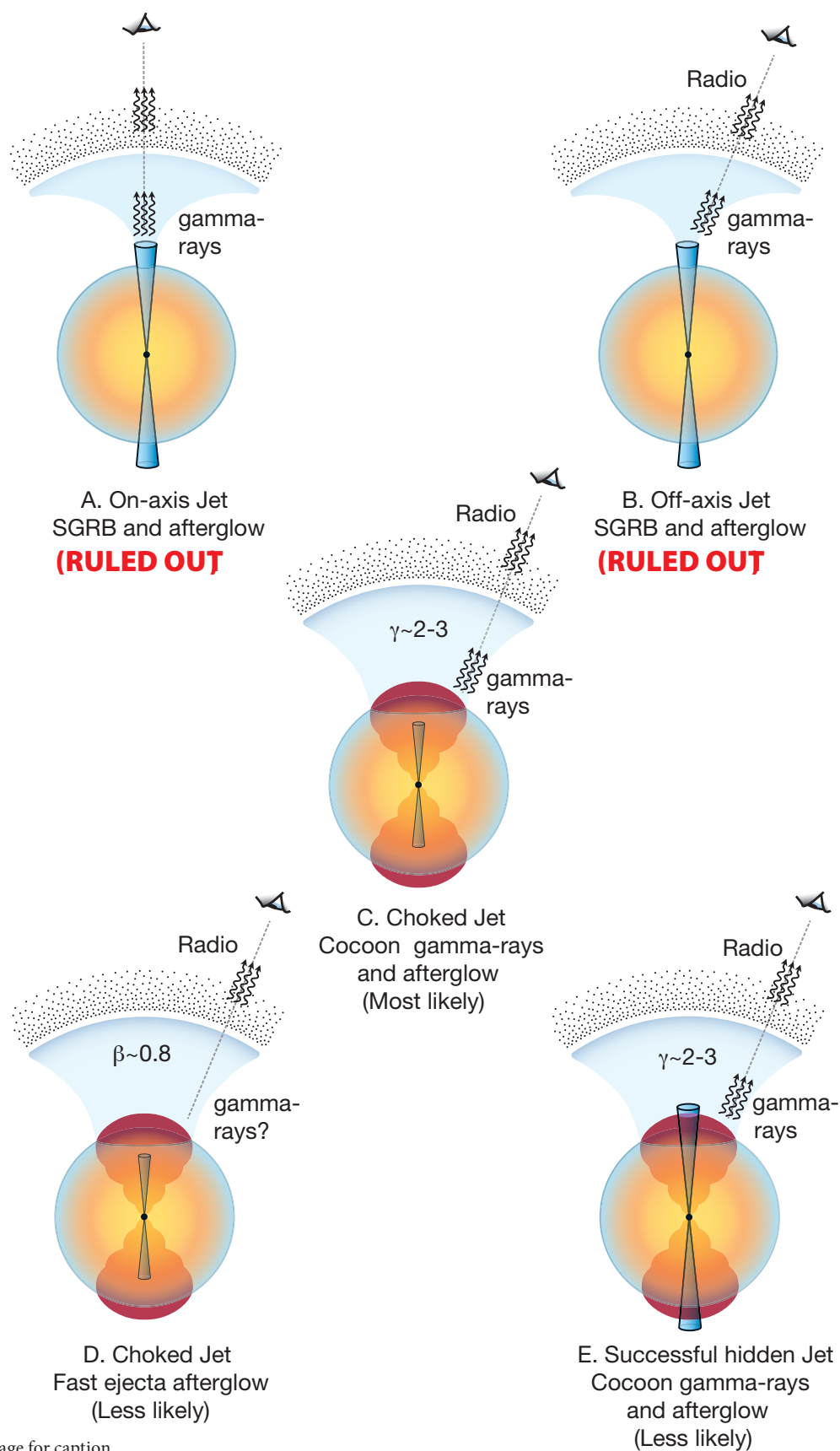
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to K.P.M. ([kunal@astro.caltech.edu](mailto:kunal@astro.caltech.edu)).

**Reviewer Information** *Nature* thanks S. Chatterjee and R. Wijers for their contribution to the peer review of this work.



**Figure 1 | The radio light curve of GW170817.** Panel (a): The flux densities corresponding to the detections (markers with  $1\sigma$  error bars; some data points have errors smaller than the size of the marker) and upper limits (markers with downward-pointing arrows) of GW170817 at frequencies ranging from 0.6–15 GHz between day 16 and day 107 post-merger (ref. 12 and Extended Data Table 1). Panel (b): Same as the panel (a) but with flux densities corrected for the spectral index  $\alpha = -0.61$  (see Methods) and early-time, non-constraining, upper limits removed.

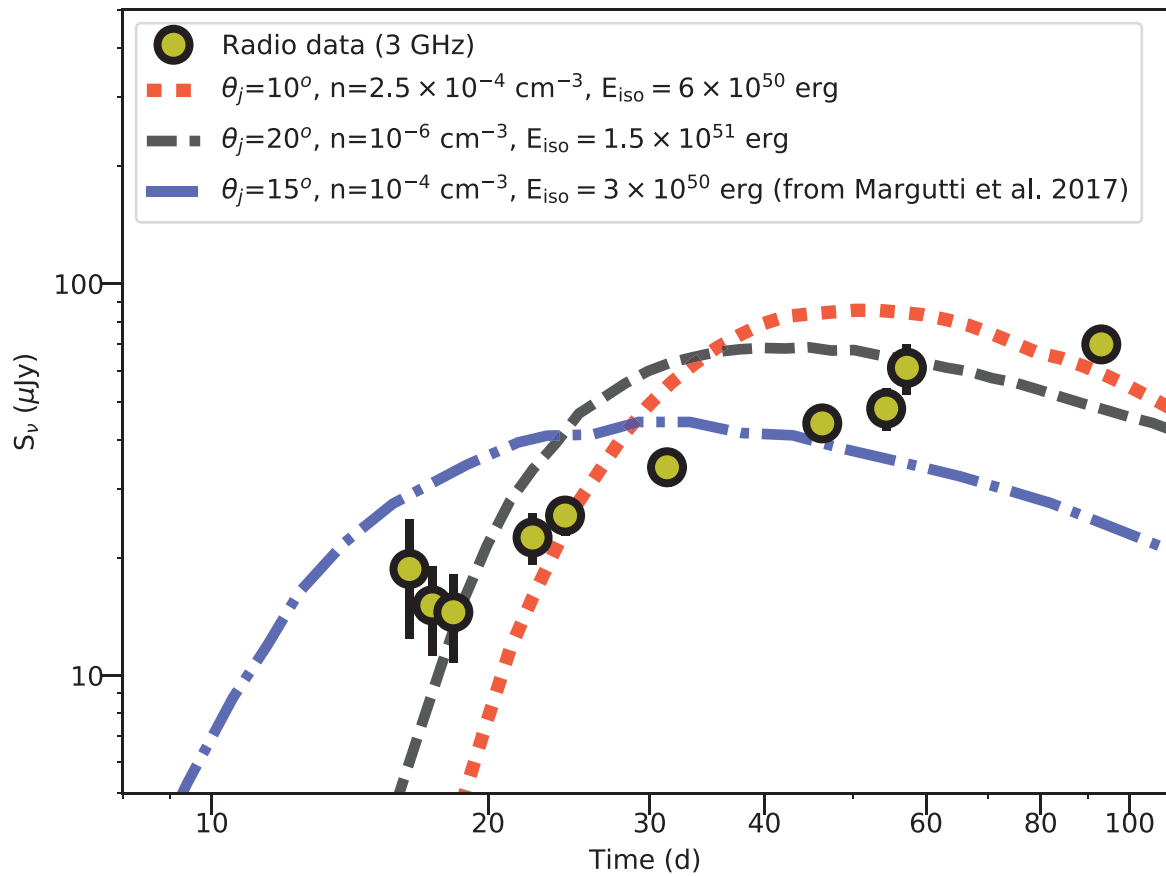
The fit to the light curve with the temporal index  $\delta = 0.78$  (see Methods) is shown as a red line and the uncertainty in  $\delta$  ( $\pm 0.05$ ) as the red shaded region. Panel (c): Residual plot after correcting for the spectral and temporal variations. The observing frequencies are color coded according to the colorbar displayed at the right (black for  $\leq 1$  GHz and yellow for  $\geq 10$  GHz). The marker shapes denote measurements from different telescopes.



**Figure 2** | See next page for caption.

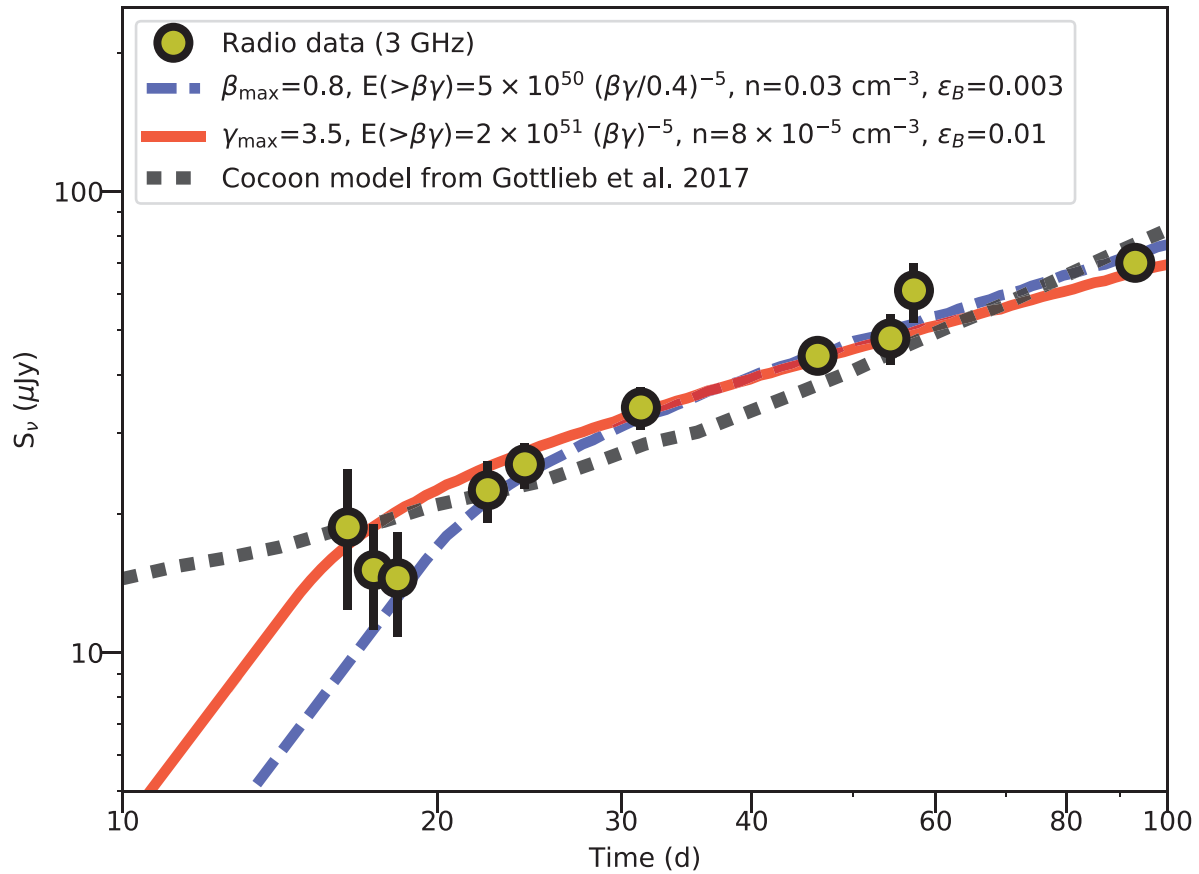
**Figure 2 | Schematic illustration of the various possible jet and dynamical ejecta scenarios in GW170817.** A) A jet seen on-axis, generating both the low-luminosity gamma-rays and the observed radio afterglow. This scenario cannot explain the late rise of the radio emission. It is also unable to explain<sup>11</sup> how a low-luminosity jet penetrates the ejecta. It is therefore ruled out. B) A regular (luminous) SGRB jet seen off-axis, producing the gamma-rays and the radio. The continuous moderate rise in the radio light curve rules out this scenario. C) A choked jet giving rise to a mildly relativistic ( $\gamma \sim 2\text{--}3$ ) cocoon which generates the gamma-rays and the radio waves via on-axis emission. This is the model that is most consistent with the observational data. It accounts for the observed gamma-rays, X-rays (possibly also the ultraviolet and optical emission) and the radio emission, and provides a natural explanation for the lack of an off-axis jet signature in the radio. D) The fast velocity tail ( $\beta \sim 0.8\text{--}0.6c$ , i.e.  $\gamma \sim 1.67\text{--}1.25$ ) of the ejecta produces the radio emission. In this case, the jet must be choked (otherwise its off-axis emission should

have been seen). While the radio emission is fully consistent with this scenario, the energy deposited in faster ejecta ( $\gamma \sim 2\text{--}3$ ) must be very low. In this scenario, the source of the observed gamma-rays remains unclear. E) A successful jet that drives a cocoon but does not have a clear signature in the radio. The cocoon generates the gamma-rays and the radio emission, and outshines the jet at all wavelengths. This scenario is less likely based on theoretical considerations, which suggest that the jet and the cocoon should have comparable energies, in which case the jet signature would have been observed in the radio band. This scenario can also be visualized as a “structured” jet, having a relativistic narrow core surrounded by a mildly relativistic wide-angle outflow, in which an off-axis observer does not see any signature of the core. The relativistic core could have produced a regular SGRB for an observer located along the axis of the jet. Such a jet, if it exists, could be too weak (made a sub-dominant contribution to the radio light curve early on) or too strong (such that its radio and X-ray signatures will be observed in the future; see Methods).



**Figure 3 | Off-axis jet models.** Synthetic light curves with a range of jet opening angles  $\theta_j$ , isotropic-equivalent energy  $E_{\text{iso}}$ , and the ISM density  $n$  (see Methods) overplotted on the 3 GHz light curve (error bars are  $1\sigma$ ; ref. 12 and Extended Data Table 1). The overall shape of the light curve remains unchanged even after changing these parameters. We have considered a wide range of parameters in the phase space of off-axis models (including unlikely scenarios like  $n = 10^{-6} \text{ cm}^{-3}$ ; see Methods);

none of the models give a good fit to the observational data, and hence we rule out the classical off-axis jet scenario as a viable explanation for the radio afterglow. The dashed black and dotted red curves are calculated using the codes described in the Methods. The dashed-dot blue curve is taken from figure 3 of ref. 4 (scaled to 3 GHz using  $\alpha = -0.6$ ). All off-axis models assume  $\theta_{\text{obs}} = 26 \text{ deg}$ ,  $\epsilon_e = 0.1$ ,  $\epsilon_B = 0.01$  and  $p = 2.2$ . (see main text and Methods).



**Figure 4 | Quasi-spherical ejecta models.** Radio light curves arising from quasi-spherical ejecta with velocity gradients, overplotted on the 3 GHz data spanning days 16–93 post-merger (filled yellow circles; error bars are  $1\sigma$ ; ref. 12 and Extended Data Table 1). The solid red and dashed blue light curves represent power law models with maximum Lorentz factors  $\gamma=3.5$  and  $\gamma=1.67$  respectively (i.e. maximum  $\beta=v/c=0.96$  and  $0.8$  respectively). These curves approximately correspond to the cocoon and dynamical ejecta, respectively. The shallow rise of the radio light

curve is consistent with a profile  $E(>\beta\gamma) \propto (\beta\gamma)^{-5}$ . For  $n \sim 0.03 \text{ cm}^{-3}$ , the observed radio emission at 93 days is produced by an ejecta component with a velocity of  $\sim 0.6c$  and kinetic energy of  $\sim 10^{49}$  erg. For a lower ISM density,  $\sim 10^{-4} \text{ cm}^{-3}$ , the radio emission at 93 days is produced by a component with a velocity of  $0.9c$  and energy  $10^{50}$  erg. Parameters  $\epsilon_e=0.1$  and  $p=2.2$  are used for both models. Also shown for reference is the cocoon model light curve (dotted black curve) taken from ref. 14, where parameter values  $n=1.3 \times 10^{-4} \text{ cm}^{-3}$ ,  $\epsilon_B=0.01$ ,  $\epsilon_e=0.1$  and  $p=2.1$  are used.

## METHODS

**1. Radio Data Analysis.** *VLA.* Radio observations of the GW170817 field were carried out with the Karl G. Jansky Very Large Array in its B configuration, under a Director Discretionary Time (DDT) program (VLA/17B-397; PI: K. Mooley). All observations were carried out with the Wideband Interferometric Digital Architecture (WIDAR) correlator in multiple bands including L-band (nominal center frequency of 1.5 GHz, with a bandwidth of 1 GHz), S-band (nominal center frequency of 3 GHz, with a bandwidth of 2 GHz), and C-band (nominal center frequency of 6 GHz, with a bandwidth of 4 GHz). We used QSO J1248-1959 (L-band and S-band) and QSO J1258-2219 (C-band) as our phase calibrator sources, and 3C 286 or 3C 147 as flux density and bandpass calibrators. The data were calibrated and flagged for radio frequency interference (RFI) using the VLA automated calibration pipeline which runs in the Common Astronomy Software Applications package (CASA<sup>25</sup>). We manually removed further RFI, wherever necessary, after calibration. Images of the observed field were formed using the CLEAN algorithm (with the “psfmode” parameter set to Hogbom<sup>26</sup>), which we ran in the interactive mode. The results of our VLA follow-up campaign of GW170817 are reported in Extended Data Table 1, and the image cutouts are shown in Extended Data Figure 1. The flux densities were measured at the Gaia/HST position<sup>27</sup>. Flux density measurement uncertainties denote the local root-mean-square (rms) noise. An additional 5% fractional error on the measured flux density is expected due to inaccuracies in the flux density calibration. For non-detections, upper-limits are calculated as three times the local rms noise in the image.

*ATCA.* We observed GW170817 on 2017 November 01, November 18 and December 02 using the Australia Telescope Compact Array (ATCA) under a target of opportunity program (CX391; PI: T. Murphy). During these observations the array was in configurations 6A, 1.5C and 6C respectively. We observed using two 2 GHz frequency bands with central frequencies of 5.5 and 9.0 GHz. For both epochs, the flux scale and bandpass response were determined using the ATCA primary calibrator PKS B1934-638, and observations of QSO B1245-197 were used to calibrate the complex gains. The visibility data were reduced using the standard routines in the MIRIAD environment<sup>28</sup>. The calibrated visibility data were split into the separate bands (5.5 GHz and 9.0 GHz), averaged to 32 MHz channels, and imported into DIFMAP<sup>29</sup>. Bright field sources were modeled separately for each band using the visibility data and a combination of point-source and Gaussian components with power-law spectra. With the field sources modelled and subtracted from the visibility data, the dominant emission in the residual image was from GW170817. Restored images for each band were generated by convolving the model components with the restoring beam, adding the residual map and then averaged to form a wide-band image. Image-based Gaussian fitting for an unresolved source was performed in the region of GW170817, leaving the flux density and source position unconstrained. The source position from the fitting agrees with the Gaia/HST position<sup>27</sup> of GW170817. The measured radio flux densities in the combined images are reported in the Extended Data Table 1, and the image cutouts are shown in Extended Data Figure 1.

*GMRT.* We carried out observations of the GW170817 field with the upgraded Giant Meterwave Radio Telescope (uGMRT) at 700 MHz under a DDT program (DDTB288; PI: K. De). All observations were carried out with 400 MHz bandwidth centered at 750 MHz using the non-polar continuum interferometric mode of the GMRT Wideband Backend (GWB<sup>30</sup>). Pointings were centered at the location of the optical transient. 3C 286 was used as the absolute flux scale and bandpass calibrators, while phase calibration was done with the sources J1248-199 (for the 2017 September 16 observation) and 3C 283 (for all other observations). These data were calibrated and RFI flagged using a custom-developed CASA pipeline. The data were then imaged interactively with the CASA task CLEAN, incorporating a few iterations of phase-only self-calibration by building a model for bright sources in the field with each iteration. The GMRT flux density measurements at the Gaia/HST position<sup>27</sup> are reported in the Extended Data Table 1. The image cutouts are shown in Extended Data Figure 1.

**1.1 Radio Data Power-law Fit.** We carried out a least-squares fit to the assembled radio data as a function of time and frequency, using a two-dimensional power-law model:

$$S(\nu, t) = S_0(\nu/\nu_0)^\alpha (t/t_0)^\delta$$

The fit results are shown in Extended Data Figure 2, where we find good results for  $\alpha = -0.61 \pm 0.05$ ,  $\delta = 0.78 \pm 0.05$ ,  $S_0 = 13.1 \pm 0.4 \mu\text{Jy}$ ,  $\nu_0 = 3 \text{ GHz}$  and  $t_0 = 10 \text{ d}$ . The fit has  $\chi^2 = 42.3$  for 44 degrees-of-freedom, although there are only 27 detections among the 47 data-points.

**1.2 Multi-epoch radio spectra.** In Extended Data Figure 3 we show the radio continuum spectra obtained at different epochs. All epochs are individually consistent with the spectral index  $\alpha = -0.61$  within 1 $\sigma$ .

**2. Model Descriptions.** **2.1 Off-axis afterglows.** The radio light curves were calculated using two independent semi-analytic codes<sup>31,32</sup>, which are based on similar approximations. Both codes were compared to, and have been found to be largely consistent with, the light curves produced by the BOXFIT code<sup>33</sup>. In short, both codes approximate the jetted blast wave at any time in the source-frame as a single zone emitting region which is a part of a sphere with an opening angle,  $\theta_j$ . The hydrodynamics includes the shock location and velocity, and the jet spreading. The hydrodynamic variables in the emitting region are set to their values immediately behind the shock. The emission from each location along the shock is calculated using standard afterglow theory<sup>34</sup>, where the microphysics is parameterized by the fraction of internal energy that goes to the electrons,  $\epsilon_e$ , the fraction of internal energy that goes to the magnetic field,  $\epsilon_B$ , and the power-law index of the electron distribution. The code calculates the rest frame emissivity at any time and any location along the shock and the specific flux observed at a given viewing angle at a given time and frequency is then found by integrating the contribution over equal-arrival-time surfaces, with a proper boost to the observer frame.

**2.2 Quasi-spherical ejecta.** Radio light curves arising from quasi-spherical outflows, e.g., a cocoon and the tail of the dynamical ejecta, are approximately described by a model with a single one-dimensional velocity profile:  $E(>\beta\gamma) \propto (\beta\gamma)^{-k}$ , where  $\beta$  is a velocity in units of the speed of light and  $\gamma$  is a Lorentz factor. The slope of the observed radio light curve is consistently explained with  $k = 5$ . The light curves are calculated using the same codes as in section 2.1. In Figure 4, we show two cases: (1) a cocoon model,  $E(>\beta\gamma) = 2 \times 10^{51} (\beta\gamma)^{-5} \text{ erg}$  with a maximum Lorentz factor of 3.5,  $n = 8 \times 10^{-5} \text{ cm}^{-3}$ , and  $\epsilon_B = 0.01$ , and (2) a dynamical ejecta model,  $E(>\beta\gamma) = 5 \times 10^{50} (\beta\gamma/0.4)^{-5} \text{ erg}$  with a maximum velocity of 0.8c,  $n = 0.03 \text{ cm}^{-3}$ , and  $\epsilon_B = 0.003$ . This velocity profile of the dynamical ejecta contains a larger mass traveling faster than 0.6c by a factor of  $\sim 5$  compared with that found in general relativistic numerical simulations<sup>20,21</sup>. The small amount of mass ejected at these high velocities is plausible since the simulations are affected by finite resolution and artificial atmosphere. In addition, Figure 4 shows a prediction from the full 2D simulation of a choked jet and the resulting cocoon presented in ref. 9. The light curve is taken from figure 4 of ref. 14 without any attempt to fit the radio data that was added since it was published. A more detailed publication reporting the full set of 2D simulations is in preparation. Finally, an upper limit on the ISM density<sup>12</sup> of  $0.04 \text{ cm}^{-3}$  suggests that the ejecta contains a fast moving component with  $v \gtrsim 0.6c$ . For all the models shown in Figure 4, the mass of the ejecta that produces the radio signal up to 93 days is only  $\sim 10^{-5} M_\odot$ . This velocity is faster, and the mass is much lower, than those inferred from the kilonova emission<sup>35</sup>. We note that kilonova ejecta will produce observable radio signals on a timescale of years.

**3. Hiding an off-axis jet.** Hiding a luminous off-axis jet (of the type seen in regular SGRBs), given the radio data, is not trivial. First, the jet emission peaks once its Lorentz factor drops to  $\sim 1/(\theta_{\text{obs}} - \theta_j)$ , where  $\theta_{\text{obs}}$  is the viewing angle with respect to the jet axis and  $\theta_j$  is the jet opening angle. Thus, emission from a jet that points only slightly away from us ( $< 10$  degrees), will peak when its Lorentz factor is high ( $\gtrsim 6$ ). Since the flux in the radio at a given time is extremely sensitive to the blast wave Lorentz factor (roughly as  $\gamma^{10}$ ) a jet at that angle will be much brighter than any on-axis mildly relativistic outflow around the peak, even if the outflow carried much more energy than the jet. Therefore, a hidden jet must be far away from the line-of sight, namely  $\theta_{\text{obs}} - \theta_j \gtrsim 10$  degrees. At such angle, any gamma-ray signal produced by a relativistic jet will be too faint to explain the observed gamma-ray signal<sup>11</sup>. Thus, while our previous radio observations strongly disfavored a regular SGRB seen off-axis as the origin of the gamma-rays<sup>12</sup> (scenario B in Figure 2), the additional observations presented here practically rule this out.

The extreme dependence of the radio flux density on the blast wave Lorentz factor also implies that, for reasonable parameters also at  $\theta_{\text{obs}} - \theta_j \gtrsim 10$  degrees, off-axis jet emission will outshine a blast wave driven by material with  $\beta \sim 0.8$  ( $\gamma \sim 1.67$ ). Thus, the radio emission from an off-axis jet may remain undetected only if the observed emission is dominated by an on-axis material with  $\gamma \sim 3$ , which is most likely a cocoon. In that case, a jet that is far from the line of sight may be hidden in two ways, either by being significantly less energetic than the on-axis outflow or, surprisingly, by being significantly more energetic (scenario E in Figure 2).

In the latter case the jet emission will not appear in the radio data available so far if it is so energetic that its Lorentz factor at day 93 is still significantly larger than  $\theta_{\text{obs}} - \theta_j$ . For example, a 10 degree jet with an isotropic equivalent energy of  $10^{52} \text{ erg}$ , that propagates in circum-merger density of  $10^{-4} \text{ cm}^{-3}$  and observed at an angle of 30 degrees, peaks after 200 days and its brightness is comparable to the observed data only around day 90 ( $\epsilon_B = 0.01$ ,  $\epsilon_e = 0.1$ ). While we cannot rule out this option, the extreme jet energies make it unlikely, but if this is the case then we will see the jet contribution in the future.

The other possibility, that the jet is less energetic than the on-axis outflow (again scenario E in Figure 2), cannot be tested observationally. However, it is unlikely

based on theoretical considerations. The energy of the cocoon is distributed over a large range of velocities. Thus, the energy of the mildly relativistic ejecta ( $\gamma \sim 3$ ) is expected to be only a small fraction of the total cocoon energy<sup>9</sup>. Moreover, observationally we see that the energy carried by slower moving on-axis material is at least a factor of 10 larger than energy carried by high velocity on-axis material. Now, the ratio between the total energy in the cocoon and the energy in the jet depends on the ratio between the time spent by the jet in the ejecta before it breaks out and the time over which the jet launching continues after the breakout takes place. The engine that launches the jet is not affected by the propagation of the jet though the ejecta and is causally disconnected from the jet head, if and when it breaks out of the ejecta. Therefore, there is no reason for the engine to stop upon breakout and without fine tuning. If the jet breaks out successfully the launching of the jet is expected to continue over a time that is comparable to or larger than the time it takes for the jet to cross the ejecta. As a result, the energy in the jet is expected to be comparable or larger than that in the cocoon. Thus, it is highly unlikely that the jet is less energetic than the fastest cocoon material, which as noted above carries only a small fraction of the total cocoon energy.

We therefore conclude that there are no probable scenarios in which a jet successfully breaks out, producing an SGRB seen by another (non-Earth) observer, and remains undetected by our radio observations. We find the case in which the jet is choked as the one that provides the best explanation to entire set of observations available to date.

**4. The origin of the gamma-rays.** Since a hidden jet cannot produce the observed gamma-rays and the rising radio emission indicates a mildly relativistic wide-angle outflow moving towards us, we can expect that this outflow is also the origin of the gamma-rays. We do not see any plausible scenario in which the kilonova ejecta can produce the gamma-rays by itself. Compactness arguments imply that this ejecta is too slow<sup>11,14</sup> and there is no natural dissipation process that can convert the kinetic energy of the ejecta to gamma-rays. The cocoon, on the other hand, can produce the gamma-rays. It has sufficient energy and its Lorentz factor is sufficiently high to avoid compactness issues, so in the presence of a dissipation mechanism it can produce the observed gamma-rays<sup>9,10,36,37</sup>. For example, a breakout of the shock driven by the cocoon through the expanding ejecta can produce the observed signal, accounting for its luminosity, duration, peak energy and spectral evolution<sup>9</sup>.

**5. Lower limit to the circum-merger density.** The mean cosmological baryon density is a function of the D/H ratio<sup>38</sup>, primordial Helium density<sup>39</sup>, cosmographic parameters<sup>40</sup> and the fraction of diffuse baryons in the IGM ( $f_{\text{IGM}}$ ) and is given<sup>41</sup> as,

$$n_{\text{H}} \sim (1.88 \times 10^{-7} \text{ cm}^{-3}) f_{\text{IGM}} (1+z)^3$$

We adopt<sup>41</sup>  $f_{\text{IGM}} = 0.7$ . At  $z \sim 0$ , a density of  $10^{-6} \text{ cm}^{-3}$  corresponds to a baryon over density  $\Delta_b = 5$ . For the Lyman-alpha forest,  $\Delta_b$  is in the range of 10-50, whereas that in condensed halos<sup>41</sup> is  $10^2 < \Delta_b < 10^4$ . Thus, in the case of GW170817, a lower limit to the ambient density is  $2 \times 10^{-5} \text{ cm}^{-3}$  and a typical value<sup>42</sup> would be  $\sim 10^{-4} \text{ cm}^{-3}$ .

**6. Radio-X-ray comparison.** The 3 GHz flux density measured<sup>12</sup> on 2017 September 03.9 is  $15 \pm 4 \mu\text{Jy}$ . Scaling the X-ray fluxes given in ref. 6 (reported in the energy range 0.3-8 keV) to the values reported in ref. 5 (0.3-10 keV), we estimate the X-ray flux on 2017 September 02.2 as  $5.5 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ , with a  $1\sigma$  uncertainty of  $\sim 1.5 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ . We use this information (X-ray flux density is  $0.23 \pm 0.06 \text{ nJy}$  at a nominal center frequency of  $4 \times 10^{17} \text{ Hz}$ ) to calculate the spectral index between the radio and X-ray frequencies as  $-0.60 \pm 0.03$ . This is consistent with our estimated value of the radio-only spectral index,  $-0.61 \pm 0.05$ , within  $1\sigma$ . Therefore the radio emission and X-rays likely originate from the same source, and the cooling frequency  $\sim 16$  days post-merger is well above the soft X-ray frequencies. Extended Data Figure 4 shows a panchromatic spectrum between the radio and X-ray frequencies. Ultraviolet and near-infrared data are also plotted for comparison. Although the early-time emission in the ultraviolet, optical and infrared frequencies was dominated by thermal emission, at late times there should be a significant synchrotron component. Using the temporal and spectral indices estimated for the radio-only data (earlier in the Methods section), and assuming the cooling break remains beyond  $10^{18} \text{ Hz}$ , we can predict the X-ray flux densities between 0.3-2.2 nJy (flux between  $7 \times 10^{-15}$  to  $52 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$  in the 0.3-10 keV band) on 2017 November 18 (and also for the Chandra observation

on December 03-06). We note that, subsequent to the submission of this paper, the X-ray observations took place and confirmed our prediction. We estimate the synchrotron cooling frequency as:

For  $\gamma > 1$  (as expected for cocoon):

$$\nu_c \approx 7 \times 10^{19} \text{ Hz} \left( \frac{\gamma}{2} \right)^{-4} \left( \frac{n}{10^{-4} \text{ cm}^{-3}} \right)^{-\frac{3}{2}} \left( \frac{\epsilon_B}{0.01} \right)^{-\frac{3}{2}} \left( \frac{t}{100 \text{ days}} \right)^{-2}$$

For  $\beta < 1$  (i.e.  $\gamma \sim 1$ ; as expected for the dynamical ejecta tail):

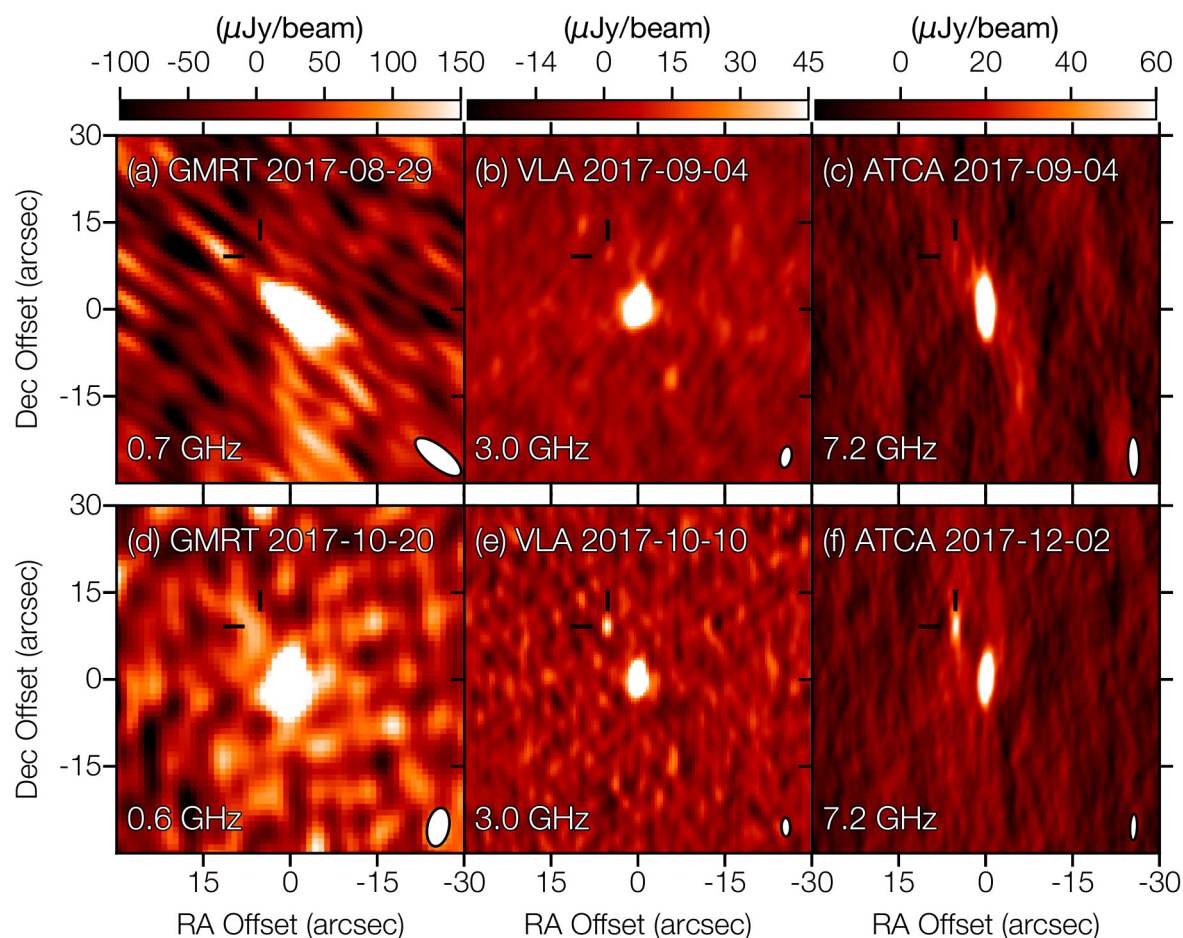
$$\nu_c \approx 2 \times 10^{18} \text{ Hz} \left( \frac{\beta}{0.6} \right)^{-3} \left( \frac{n}{0.03 \text{ cm}^{-3}} \right)^{-\frac{3}{2}} \left( \frac{\epsilon_B}{0.003} \right)^{-\frac{3}{2}} \left( \frac{t}{100 \text{ days}} \right)^{-2}$$

We see that the cooling frequency at day  $\sim 16$  post-merger is much larger than  $10^{18} \text{ Hz}$ , while beyond  $\sim 10^2$ - $10^3$  days post-merger this break should be seen moving towards lower frequencies within the electromagnetic spectrum.

**Data availability.** All relevant data are available from the corresponding author on request. Data presented in Figure 1 are included in Extended Data Table 1.

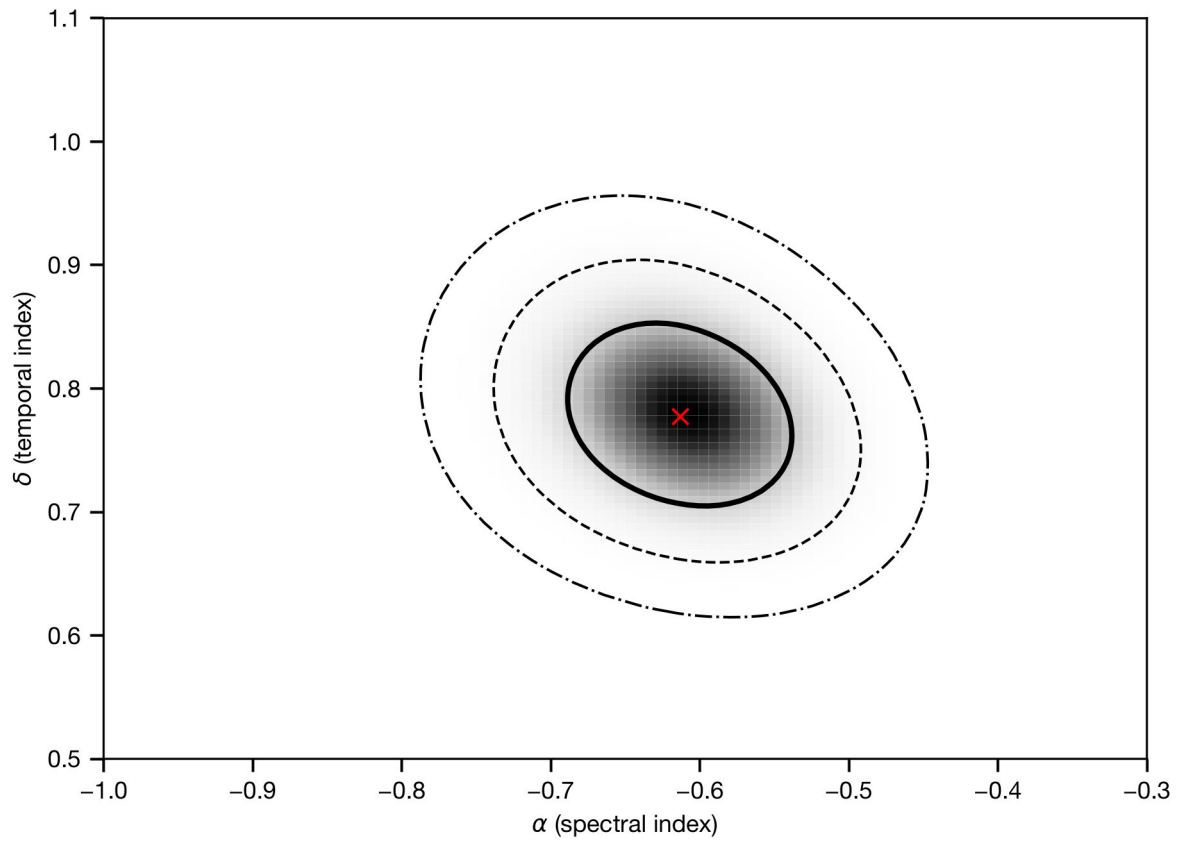
**Code availability.** The codes used for generating the synthetic radio light curves are currently being readied for public release (publication in preparation). Radio data processing software: CASA, MIRIAD, DIFMAP.

25. McMullin, J. P., Waters, B., Schiebel, D., et al. *CASA Architecture and Applications. Astronomical Data Analysis Software and Systems XVI ASP Conf. Series*, Edited by R.A. Shaw, H.E. Payne, and J.J.E. Hayes, **376**, 127–130 (2007)
26. Högbom, J. A., Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astron. Astrophys. Supp.*, **15**, 417–426 (1974)
27. Adams, S. M., Kasliwal, M. M. & Blagorodnova, N., Gamma-Ray Coordinates Network Circular 21816 (2017)
28. Sault, R. J., Teuben, P. J. & Wright, M. C. H., *A Retrospective View of MIRIAD. Astronomical Data Analysis Software and Systems IV, ASP Conf. Series*, Edited by R.A. Shaw, H.E. Payne, and J.J.E. Hayes, **77**, 433–436 (1995)
29. M. C. Shepherd, Difmap: an Interactive Program for Synthesis Imaging. *Astronomical Data Analysis Software and Systems VI ASP Conf. Series*, Edited by G. Hunt & H.E. Payne, **125**, 77–84 (1997)
30. Reddy, S. H., Kudale, S., Gokhale, U., et al. A Wideband Digital Back-End for the Upgraded GMRT. *J. Astron. Instr.*, **6**, 1641011 (2017)
31. Soderberg, A. M., Nakar, E., Berger, E. & Kulkarni, S. R., Late-Time Radio Observations of 68 Type Ibc Supernovae: Strong Constraints on Off-Axis Gamma-Ray Bursts. *Astrophys. J.*, **638**, 930–937 (2006)
32. Hotokezaka, K. & Piran, T., Mass ejection from neutron star mergers: different components and expected radio signals. *Mon. Not. R. Astron. Soc.*, **450**, 1430–1440 (2015)
33. van Eerten, H. J. & MacFadyen, A. I., Observational Implications of Gamma-Ray Burst Afterglow Jet Simulations and Numerical Light Curve Calculations. *Astrophys. J.*, **751**, 155 (2012)
34. Sari, R., Piran, T. & Narayan, R., Spectra and Light Curves of Gamma-Ray Burst Afterglows. *Astrophys. J. Lett.*, **497**, L17–L20 (1998)
35. Villar, V. A., Guillochon, J., Berger, E., et al. The Complete Ultraviolet, Optical, and Near-Infrared Light Curves of the Kilonova Associated with the Binary Neutron Star Merger GW170817: Homogenized Data Set, Analytic Models, and Physical Implications. eprint arXiv:1710.11576 (2017)
36. Nakar, E. & Piran, T. The Observable Signatures of GRB Cocoon. *Astrophys. J.*, **834**, 28 (2017)
37. Lazzati, D., Lopez-Camara, D., Cantiello, M., Morsony, B. J., Perna, R. & Workman, J. C. Off-axis Prompt X-Ray Transients from the Cocoon of Short Gamma-Ray Bursts. *Astrophys. J.*, **848**, 6 (2017)
38. Cooke, R. J., Pettini, M., Nollett, K. M. & Jorgenson, R. The Primordial Deuterium Abundance of the Most Metal-poor Damped Lyman- $\alpha$  System. *Astrophys. J.*, **830**, 148 (2016)
39. Aver, E., Olive, K. A., Porter, R. L. & Skillman, E. D., The primordial helium abundance from updated emissivities. *JCAP*, **11**, 017 (2013)
40. Planck collaboration 2016. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, **594**, 13 (2016)
41. Shull, M. J., Smith, B. D. & Danforth, C. W. The Baryon Census in a Multiphase Interstellar Medium: 30% of the Baryons May Still be Missing. *Astrophys. J.*, **759**, 23 (2012)
42. J. Tumlinson, M. S. Peebles & J. K. Werk. The Circumgalactic Medium. *Ann. Rev. Astron. Astrophys.*, **55**, 389 (2017)



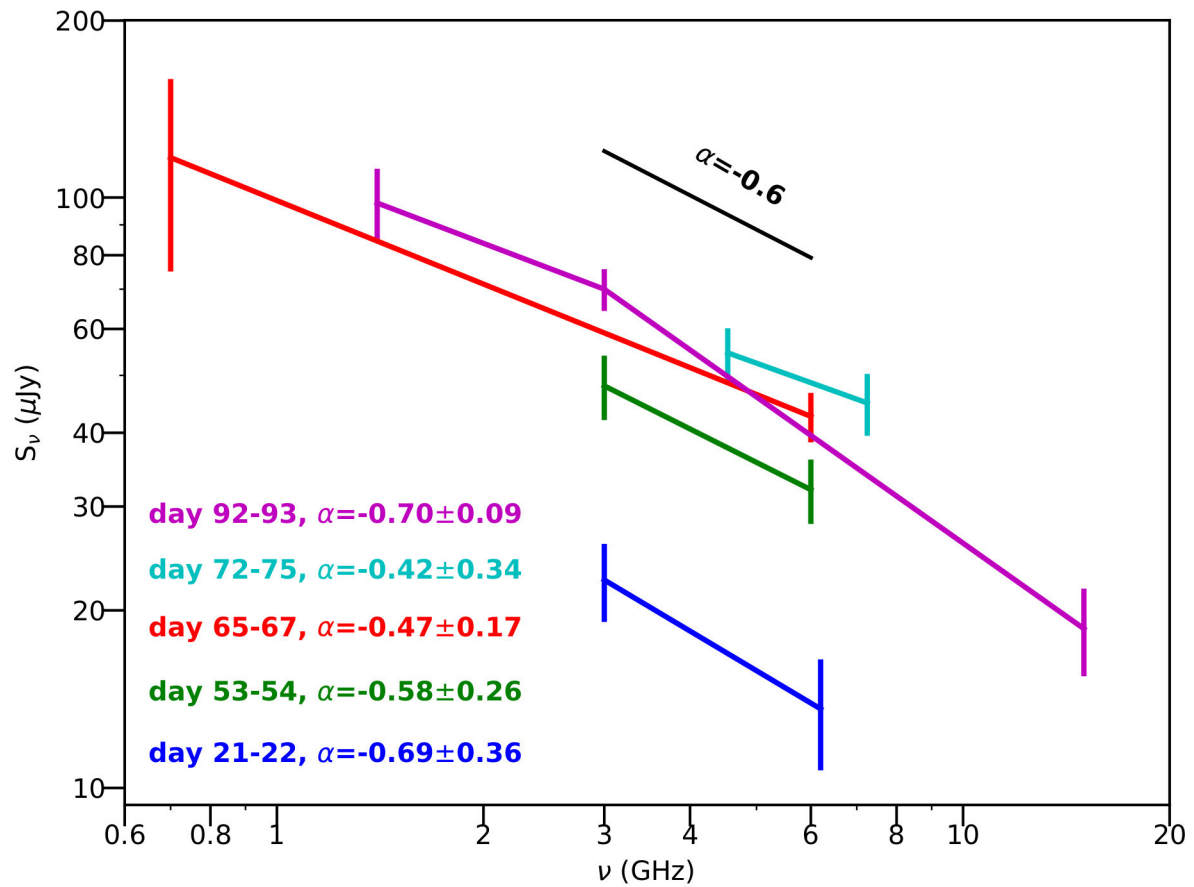
**Extended Data Figure 1 | GW170817 radio image cutouts.** Image cutouts ( $30'' \times 30''$ ) from the upgraded GMRT, the VLA and the ATCA centred on the NGC 4993. The position of GW170817 is marked by two black lines. Panels (a), (b) and (c) show images from August-September 2017, using

the data reported in ref. 12. Panels (d), (e) and (f) show recent data, from October 2017. The flux density scale is denoted by the colorbar in each column. The synthesized beam is shown as an ellipse in the lower right corner of each image.



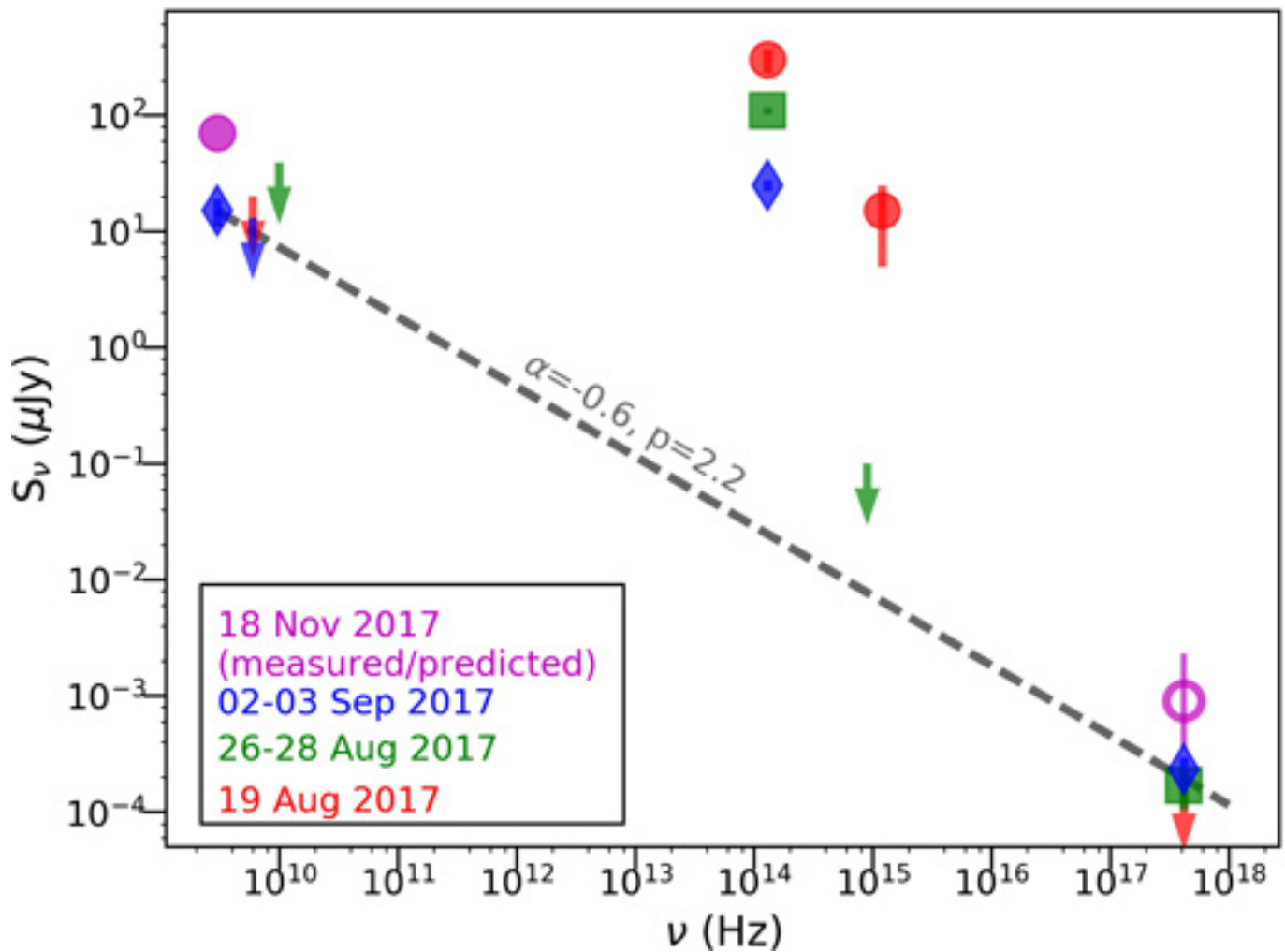
**Extended Data Figure 2 | Confidence region for the radio spectral and temporal indices.** Joint confidence contours for  $\alpha$  (the spectral power-law index) and  $\beta$  (the temporal power-law index). The contours are 1-, 2-, and

3- $\sigma$  confidence contours, and the location of the best-fit values,  $\alpha = -0.61 \pm 0.05$ ,  $\delta = 0.78 \pm 0.05$ , is indicated by the red “x” marker.



**Extended Data Figure 3 | Radio-only spectral indices of GW170817.** Radio spectral indices between 0.6–15 GHz spanning multiple epochs. The different epochs are color coded. The corresponding days post-merger

and spectral indices are given in the legend. Error bars are  $1\sigma$ . The joint analysis of all radio data (see text in Methods section) implies  $\alpha = -0.61 \pm 0.05$ .



**Extended Data Figure 4 | Comparison between the radio and X-ray flux densities of GW170817.** The comparison of the X-ray data with the radio upper limits (arrows) and detections (filled circles) at different epochs. Error bars are  $1\sigma$ . The epochs: 2017 August 19, August 26–28, September 02–03 and November 18 (2 days, ~10 days, ~15 days and 93 days post-merger respectively) are color coded (the epoch is given in the legend to the upper-right corner) and marked with different symbols. The spectral index ( $\alpha$ ) and corresponding electron power law index ( $p$ ; assuming cooling frequency is beyond  $10^{18}$  Hz, as expected for a mildly relativistic outflow) between 3 GHz and  $10^{18}$  Hz as derived from the September 02–03

data ( $\alpha = -0.60 \pm 0.03$  and  $p = 2.20 \pm 0.06$ ) are consistent with the radio-only spectral indices, and shown here as a dashed grey line. This indicates that the radio and X-rays originate from the same synchrotron source. The corresponding predicted soft-X-ray flux density on November 18 (between 0.3–2.2 nJy; *note: the Chandra X-ray observations from 03–06 December, reported after the submission of this paper, confirmed the prediction*) is shown as a magenta unfilled circle with an error bar. The flux densities in the ultraviolet ( $\sim 10^{15}$  Hz) and near-infrared ( $\sim 10^{14}$  Hz), dominated by thermal emission at early times, are shown for reference.

Extended Data Table 1 | Radio data for GW170817

| UT Date   | $\Delta T$<br>(d) | Telescope | $\nu$<br>(GHz) | Bandwidth<br>(GHz) | $S_\nu$<br>( $\mu\text{Jy}$ ) |
|-----------|-------------------|-----------|----------------|--------------------|-------------------------------|
| Sep 16.25 | 29.73             | GMRT      | 0.68           | 0.2                | < 246                         |
| Sep 17.84 | 31.32             | VLA       | 3              | 2                  | $34 \pm 3.6$                  |
| Sep 21.86 | 35.34             | VLA       | 1.5            | 1                  | $44 \pm 10$                   |
| Sep 25.86 | 39.34             | VLA       | 15             | 4                  | <14.4                         |
| Oct 02.79 | 46.26             | VLA       | 3              | 2                  | $44 \pm 4$                    |
| Oct 09.79 | 53.26             | VLA       | 6              | 4                  | $32 \pm 4$                    |
| Oct 10.80 | 54.27             | VLA       | 3              | 2                  | $48 \pm 6$                    |
| Oct 13.75 | 57.22             | VLA       | 3              | 2                  | $61 \pm 9$                    |
| Oct 21.67 | 65.14             | GMRT      | 0.61           | 0.4                | $117 \pm 42$                  |
| Oct 23.69 | 67.16             | VLA       | 6              | 4                  | $42.6 \pm 4.1$                |
| Oct 28.73 | 72.20             | VLA       | 4.5            | 0.5                | $54.6 \pm 5.5$                |
| Nov 01.02 | 75.49             | ATCA      | 7.25           | 4                  | $44.9 \pm 5.4$                |
| Nov 17.93 | 92.4              | ATCA      | 7.25           | 4                  | $39.6 \pm 7$                  |
| Nov 18.60 | 93.07             | VLA       | 1.6            | 1                  | $98 \pm 14$                   |
| Nov 18.66 | 93.13             | VLA       | 3              | 2                  | $70 \pm 5.7$                  |
| Nov 18.72 | 93.19             | VLA       | 15             | 4                  | $18.6 \pm 3.1$                |
| Dec 02.89 | 107.36            | ATCA      | 7.25           | 4                  | $66.5 \pm 5.6$                |

Table notes:  $\Delta T$  represents the time post-merger. The Nov 17.93 ATCA observation was affected by bad weather, and the uncertainty in the flux density is expected to be much larger than the one reported here.

# Magnetic cage and rope as the key for solar eruptions

Tahar Amari<sup>1</sup>, Aurélien Canou<sup>2</sup>, Jean-Jacques Aly<sup>3</sup>, Francois Delyon<sup>4</sup> & Frédéric Alauzet<sup>5</sup>

Solar flares are spectacular coronal events that release large amounts of energy. They are classified as either eruptive or confined<sup>1,2</sup>, depending on whether they are associated with a coronal mass ejection. Two types of model have been developed to identify the mechanism that triggers confined flares, although it has hitherto not been possible to decide between them because the magnetic field at the origin of the flares could not be determined with the required accuracy<sup>3–8</sup>. In the first type of model, the triggering is related to the topological complexity of the flaring structure, which implies the presence of magnetically singular surfaces<sup>9–11</sup>. This picture is observationally supported by the fact that radiative emission occurs near these features in many flaring regions<sup>12–17</sup>. The second type of model attributes a key role to the formation of a twisted flux rope, which becomes unstable. Its plausibility is supported by simulations<sup>18–22</sup>, by interpretations of some observations<sup>23,24</sup> and by laboratory experiments<sup>25</sup>. Here we report modelling of a confined event that uses the measured photospheric magnetic field as input. We first use a static model to compute the slowly evolving magnetic state of the corona before the eruption, and then use a dynamical model to determine the evolution during the eruption itself. We find that a magnetic flux rope must be present throughout the entire event to match the field measurements. This rope evolves slowly before saturating and suddenly erupting. Its energy is insufficient to break through the overlying field, whose lines form a confining cage, but its twist is large enough to trigger a kink instability, leading to the confined flare, as previously suggested<sup>18,19</sup>. Topology is not the main cause of the flare, but it traces out the locations of the X-ray emission. We show that a weaker magnetic cage would have produced a more energetic eruption with a coronal mass ejection, associated with a predicted energy upper bound for a given region.

It is important to study the pre-eruptive magnetic structure of the active region before a solar eruption (which must be temporally and spatially isolated) with sufficient precision to capture detailed features such as the presence of a twisted flux rope (TFR). It is equally important to have an efficient magnetohydrodynamic model that is able to describe the eruptive phase and to identify the underlying triggering mechanism—for instance, the ideal magnetohydrodynamic instability of a TFR. The problem is notoriously difficult and previous attempts have not provided a clear answer to the basic questions<sup>3–8</sup>. We have been able, however, to develop such efficient tools, and we apply them here to a specific flare-producing active region of the Sun.

Our target is the National Oceanic and Atmospheric Region (NOAA) active region AR12192, which was visible at the surface of the Sun in October 2014. This region is interesting because it exhibited huge spots, making it the largest region observed during more than two solar cycles (it was as large as Jupiter) and also the most active one of the current solar cycle<sup>26</sup> (the 24th since 1755). However, it produced neither coronal mass ejections nor solar energetic particles, which seemed puzzling<sup>4–8</sup>. But it was the site of very intense X-class flares, which were

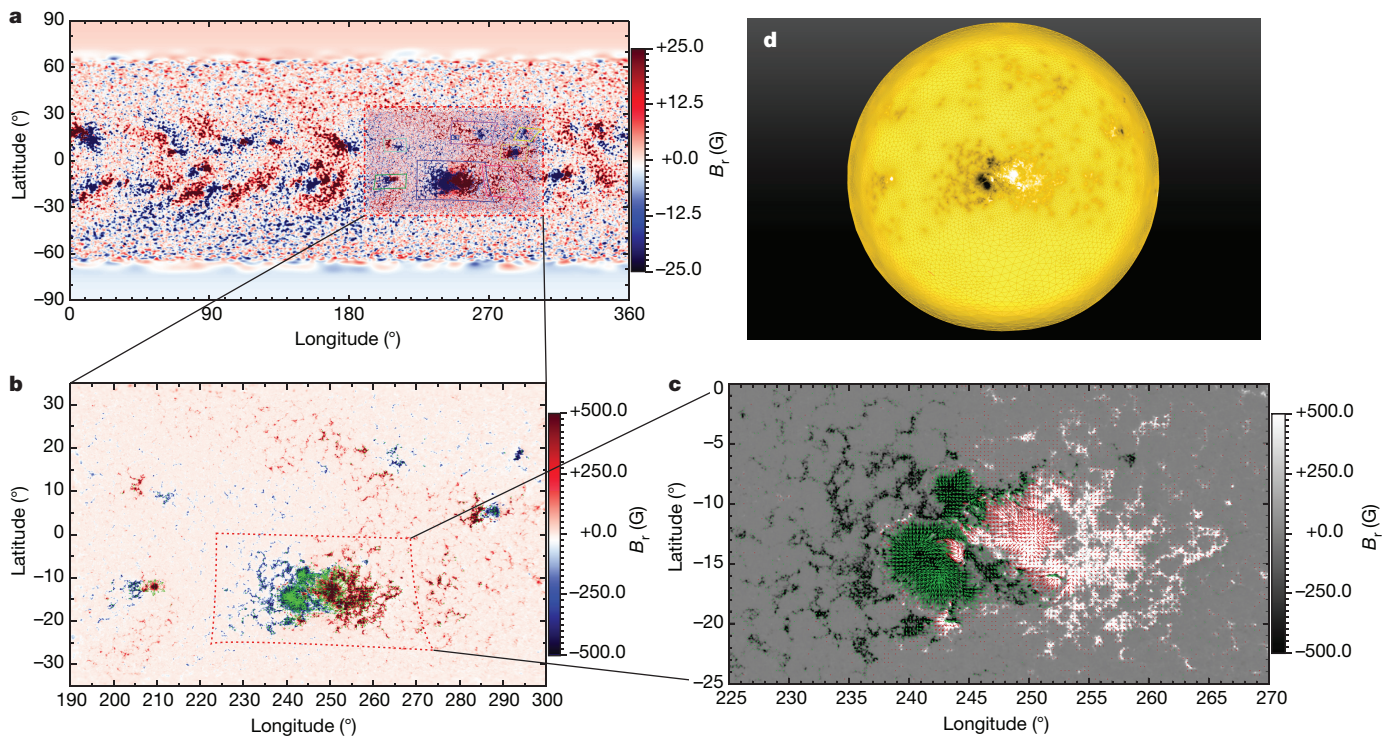
observed by the NASA Solar Dynamics Observatory (SDO) satellite and disturbed communication systems. Here we focus on the intense X3.1 flare that occurred at 21:10 UT (universal time) on 24 October 2014, a day the region was particularly active. One of the rationales for this choice is that the filtergram of the Helioseismic and Magnetic Imager (HMI) onboard SDO provided a series of high-precision measurements of the photospheric field around the time of the eruption. This prompted other groups also to model this flare, but the configurations they obtained fundamentally differ from the ones reported below. For instance, a reconstruction of the pre-eruptive magnetic structure<sup>8</sup> shows only the presence of a weakly twisted tube overlying a strongly sheared arcade, while a fully data-driven picture<sup>3</sup> (including both the pre-eruptive and eruptive phases) does not show the formation of any TFR. Therefore it is not surprising that both papers propose an interpretation of the X3.1 flare—based on a tether-cutting reconnection process—that is at odds with ours.

X-ray flux data from the Geostationary Operational Environmental Satellite (GOES) indicate that between 10:00 UT and 21:00 UT the only eruptive activity is a small C5-class flare occurring around 14:00 UT. This point was confirmed by checking with all wavelength data from the Atmospheric Imaging Assembly (AIA) instrument. To be sure to use data not affected by previous activity, however weak, we decided to restrict our attention to the quiet period running from 16:00 UT to 21:00 UT on 24 October. In a first step, during that time interval we made a series of static reconstructions of the environment of the active region at the global Sun scale. Our main tool is the code MESHMHDD<sup>27</sup>, which we fed with composite SDO/HMI data (Fig. 1) prepared as explained in Methods. MESHMHDD is a very accurate numerical code that can be used to treat both static and dynamic problems and has the characteristic of being able to achieve high resolution only where it is needed—inside the active region<sup>28</sup> at small scales. Our static reconstruction neglects the effects of the plasma and gravity forces, a common approximation that is well justified in the low corona. The computed equilibrium state is to a high degree force-free and divergence-free, which is crucial for solving the present problem and allows an accurate computation of important quantities, such as the magnetic helicity (see Methods).

Our computation of a sequence of equilibria using this method shows (Fig. 2) that the active region configuration evolves slowly, revealing the progressive appearance of a TFR. The TFR is fully formed at 21:00 UT, where it reaches its maximum size (Figs 2 and 3). It is strongly confined inside an overlying magnetic cage, whose magnetic lines are not parallel to its axis (Fig. 3c) and which plays an important part in the evolution of the configuration, in agreement with recent observations<sup>24</sup>. The magnetic twist of the TFR increases continuously (Extended Data Fig. 1). For the configuration obtained at 21:00 UT, that is, ten minutes before the major confined X3.1 flare, the twist has a value of about 2.5 turns (Extended Data Fig. 1). The structure of the electric current (and then of the force-free function  $\alpha$ )

<sup>1</sup>CPHT, Ecole Polytechnique, CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau, France. <sup>2</sup>Centre de Physique Théorique, Ecole Polytechnique, F-91128 Palaiseau Cedex, France.

<sup>3</sup>Université Paris Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, F-91191 Gif-sur-Yvette, France. <sup>4</sup>LPTMC, UMR 7600 of CNRS, Université Pierre et Marie Curie, Paris, France. <sup>5</sup>INRIA Saclay Ile-de-France, Projet Gamma 3, 1 rue Honoré d'Estienne d'Orves, 91126 Palaiseau, France.

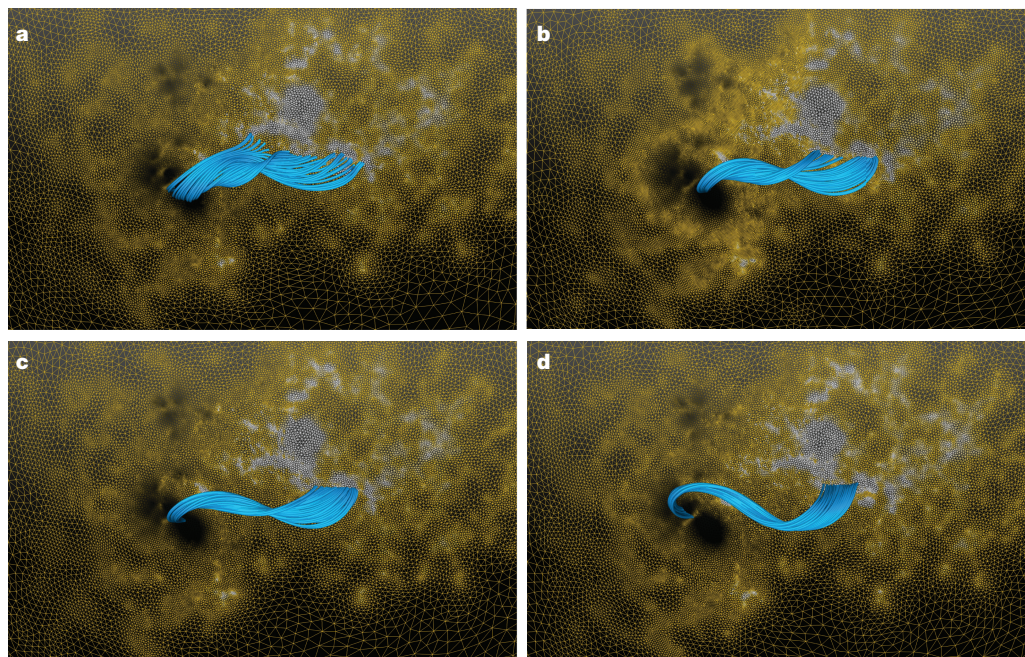


**Figure 1 | Building magnetic input for the model.** SDO/HMI high-resolution photospheric magnetic maps are used to build up adapted resolution boundary conditions.  $B_r$  denotes the radial component of the magnetic field. The composite map contains three different maps: **a**, the synoptic map of the Carrington Rotation 2156; **b**, the line-of-sight full-disk magnetogram in the Carrington heliographic coordinates (CHC); and **c**, the vector magnetogram patches from the Spaceweather Helioseismic and Magnetic Imager Active Region Patch (SHARP) of the Solar

Dynamics Observatory data series in CHC. **d**, An iterative adapted scheme (on current and magnetic field) on a triangular mesh of the photosphere results in high-resolution boundary conditions only where necessary. The coloured-outline boxes in **a** indicate the positions of the patches provided by SDO, where the vector magnetic field has been measured; the dotted-outline region enlarged in **b** corresponds to the full-disk SDO/HMI magnetogram.

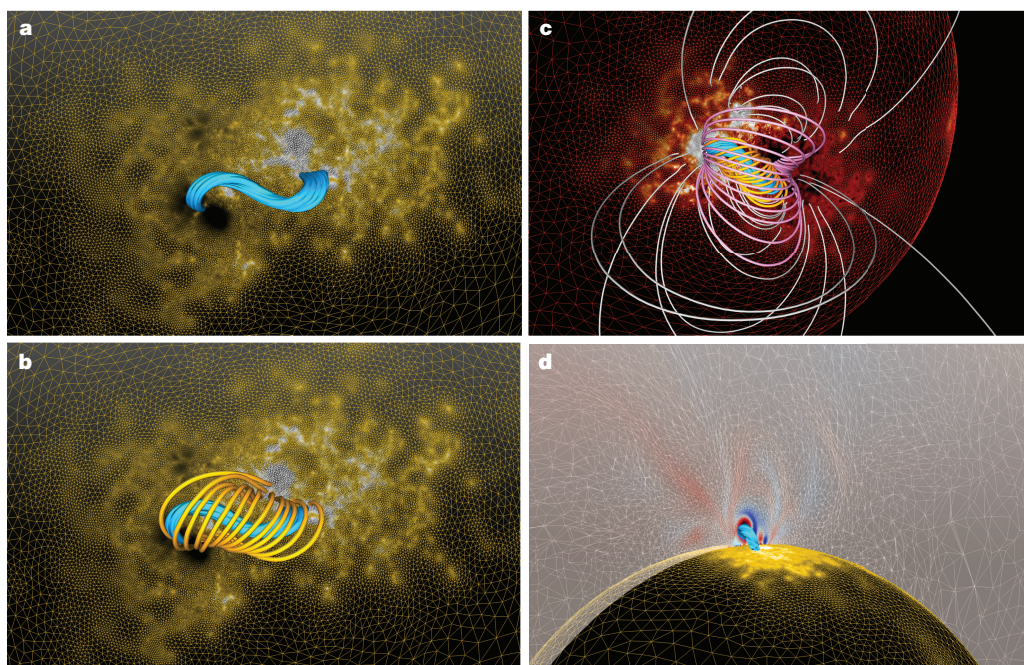
can also be determined thanks to our unstructured adapted mesh model. The current profile (Fig. 3d) appears to be complex. It contains both direct and return currents, with some of them flowing in

background twisted ropes of non-negligible twist (Extended Data Fig. 2). Like the twist, the total magnetic energy  $W$  increases during the evolution (Extended Data Fig. 1), reaching a value of  $1.8 \times 10^{34}$  erg



**Figure 2 | Magnetic field evolution before the major flare.** Selected field lines of the reconstructed magnetic configurations starting on 24 October 2014 at 16:00 UT are shown at different stages during the rising towards the

major X3.1 flare: **a**, at 16:00 UT =  $H - 5$  (five hours before the flare); **b**, at  $H - 3$ ; **c**, at  $H - 2$ ; and **d**, at hour  $H^*$ , ten minutes before the flare.



**Figure 3 | TFR in the magnetic cage before the major confined eruption.** We show characteristic properties of the configuration at H\* in Fig. 2d. **a**, A TFR covering an active region the size of Jupiter has acquired a large twist (2.5 turns). **b**, Magnetic cage covering and maintaining the TFR. **c**, Larger-scale view of the multilayer structure of the magnetic cage.

Only two layers have been plotted. **d**, View of the electric current system associated with the TFR, revealed by the mesh adaptation scheme. The colour code is symmetric, with blue (red) for negative (positive) values and saturation fixed at  $(-0.36, 0.36) \text{ Mm}^{-1}$ .

in the last computed configuration (a much smaller value was previously claimed<sup>8</sup>). If we restrict our attention to the sub-region containing the TFR, the stored magnetic energy is a factor of 1.09 above the potential energy  $W_{\text{pot}}$  of the whole region, but a factor of 1.24 above the part of  $W_{\text{pot}}$  that is contained in the sub-region.

Up to the time of the flare, the slowly evolving magnetic field with its embedded TFR stays well confined. This is in accordance with the fact that  $W$  stays below the energy  $W_{\text{sop}}$  of the associated semi-open field<sup>29</sup> (Extended Data Fig. 1; a field is expected to suffer a fast expansion and to open when  $W/W_{\text{sop}}$  approaches unity). For completeness, we have also computed the standard decay index  $n$  of the potential field associated with the pre-eruptive configuration.  $n$  is found (Extended Data Fig. 3) to remain definitely below the critical value  $n_c \approx 1.5$  that is generally considered to characterize the onset of the torus instability (this instability is thought to trigger the loss of confinement of the TFR in ejective events<sup>19</sup>). This allows us to anticipate the confined character of the flare. We point out, however, that the applicability of the general argument is somewhat uncertain in the situation we encounter here. The quoted theoretical value of  $n_c$  is computed by considering a configuration in which an idealized TFR stays in equilibrium in a background potential magnetic field, while the pre-eruptive configurations exhibit non-negligible electric currents flowing outside the TFR (Extended Data Fig. 2). Hence the background field appears to be force-free rather than current-free, which may be expected to lead to a modification of  $n_c$ . A similar point was made in a previous study of the decay index  $n$  just before the X3.1 flare<sup>8</sup>.

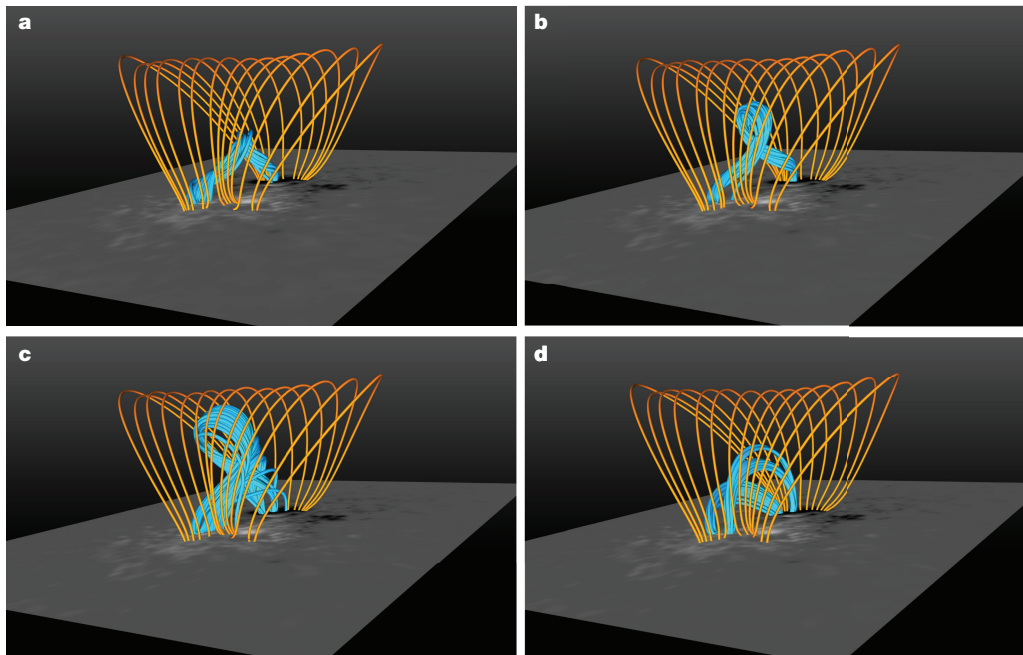
To illustrate the importance of the background currents in the determination of the actual magnetic field  $\mathbf{B}$ , we find it useful to compare its decrease with that of an auxiliary force-free magnetic field  $\mathbf{B}'$  for which the electric currents are fully confined inside a TFR that differs as little as possible from the TFR of  $\mathbf{B}$  (then  $\mathbf{B}'$  is current-free outside its TFR; see Methods). The difference between the behaviours of both force-free fields appears quite neatly (Extended Data Fig. 4), with  $\mathbf{B}$  showing a much stronger value above the central part of the TFR. As for the free energies, we find that  $W - W_{\text{pot}} \approx 19 \times 10^{32} \text{ erg}$  for  $\mathbf{B}$ , while  $W' - W_{\text{pot}} \approx 13 \times 10^{32} \text{ erg}$  for  $\mathbf{B}'$ . Then there is a huge reserve of about

$6 \times 10^{32} \text{ erg}$  in the background, which would be sufficient by itself to power an X-class flare.

The quality of our reconstructions has been checked by computing the evolution and the asymptotic convergence along the iterative scheme of two standard diagnostics measuring, respectively, the extent to which the field is force-free and divergence-free (Extended Data Fig. 5). We have also compared the pre-eruptive configuration computed at 21:00 UT with various observational data (Extended Data Figs 6 and 7).

The second step of our method consists of testing the stability of the previously obtained pre-eruptive state. Our tool here is our numerical model METEOSOL<sup>30</sup>, which is a magnetohydrodynamics code in Cartesian geometry, and we perform the computation in a numerical box that is large enough to allow for possible ejective processes<sup>29</sup>. Each of the different states computed before 21:00 UT is taken as an initial state for the model, and is made to evolve by applying, in turn, during a short time, various boundary conditions<sup>29</sup> that mimic physical processes occurring at the Sun's surface: flux cancellation, converging motions and turbulent diffusion. We found that whatever the driving process, only the configuration of 21:00 UT leaves its equilibrium state, owing to the nonlinear development of a kink instability associated with the high value of the twist (see Methods). The TFR is most clearly seen changing its original shape into the helical one that is characteristic of the kink instability (Fig. 4). At the same time, it expands slightly, making it press against the cage above, which contains some non-negligible magnetic free energy associated with the presence of electric currents and shearing of the lines (this is the reserve of about  $6 \times 10^{32} \text{ erg}$  mentioned above). This transformation of the configuration is favourable for triggering a reconnection process between the lines of the TFR and the lines forming the different layers of the cage. This leads to the conversion of some amount of magnetic energy that is tapped from both regions, and to the production of the observed flare.

No fast expansion and opening of the field are obtained in our simulations. The flare we compute is a confined one, and our results do not exhibit any disagreement with the conclusion above based on the decay index and the energy of the semi-open field. We also note that the shearing of the cage, which is made of magnetic lines whose direction



**Figure 4 | Evolution and confined instability of the TFR.** We show selected field lines of some configurations obtained during the evolution driven by small photospheric perturbations of the configuration of 24 October 2014 at 21:00 UT, taken as the initial condition. **a**, Initially

perturbed state. **b**, Some time later, the TFR experiences a kink instability. **c**, The evolution completes later. **d**, The final state of the instability, in which reconnection leads to the formation of two pieces of the TFR.

varies with altitude (Fig. 3c), may play an important part in the torus stability of the TFR, in the same way that the shearing of a field supporting plasma against gravity may stabilize Rayleigh–Taylor modes. When the kink instability develops and the TFR deforms into a helical shape, the shearing ensures that in the cage above the TFR there are always magnetic lines having a non-negligible component orthogonal to the TFR; this is the component that is important for the confinement<sup>25</sup>.

Our results show that the role of the cage, and thus of the environment, crucially affects the class of eruption—either confined or eruptive—that can be produced in an active region. Indeed, applying flux cancellation for a longer time and on a larger scale that includes the cage results in a decrease of the confinement that leads to a major disruption of the structure and to the ejection of the TFR, that is, a coronal mass ejection is produced as in our previous work<sup>29</sup> (Extended Data Fig. 9). Moreover, the maximum energy that could be stored in the configuration—which is the energy of the semi-open field—is found to be approached from below in the eruptive event (as previously found<sup>29</sup>), whereas the energy is well below it in the confined one (Extended Data Fig. 1). Thus the free energy of the semi-open field appears to give an estimate of the energy that could be released from a region by the ‘most energetic’ possible eruption.

The result previously obtained in idealistic numerical simulations<sup>18–20</sup> is thus now shown to occur on the real Sun. We conclude that the cause of the confined eruption is the instability of the TFR, and not the topological structure of the configuration. It should be noted that our model shows an interesting correlation between the location of the predicted magnetic structure and the real Sun observations during the eruption (Extended Data Fig. 8).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 21 May; accepted 20 October 2017.**

- Schrijver, J. C. *et al.* Understanding space weather to shield society: a global road map for 2015–2025 commissioned by COSPAR and ILWS. *Adv. Space Res.* **55**, 2745–2807 (2015).

- Schmieder, B., Aulanier, G. & Vrsnak, B. Flare-CME models: an observational perspective. *Sol. Phys.* **290**, 3457–3486 (2015).
- Jiang, C. *et al.* How did a major confined flare occur in super solar active region 12192? *Astrophys. J.* **828**, 62 (2016).
- Liu, R. *et al.* Structure, stability, and evolution of magnetic flux ropes from the perspective of magnetic twist. *Astrophys. J.* **818**, 148 (2016).
- Yang, K., Guo, Y. & Ding, M. D. Quantifying the topology and evolution of a magnetic flux rope associated with multi-flare activities. *Astrophys. J.* **824**, 148 (2016).
- Wang, H. *et al.* Witnessing magnetic twist with high-resolution observation from the 1.6-m new solar telescope. *Nat. Commun.* **6**, 7008 (2015).
- Sun, X. *et al.* Why is the great solar active region 12192 flare-rich but CME-poor? *Astrophys. J.* **804**, L28 (2015).
- Inoue, S., Hayashi, K. & Kusano, K. Structure and stability of magnetic fields in solar active region 12192 based on nonlinear force-free field modeling. *Astrophys. J.* **818**, 168 (2016).
- Gorbachev, V. S., Kelner, S. R., Somov, B. V. & Shvarts, A. S. A new topological approach to the question of the trigger for solar flares. *Sov. Astron.* **32**, 308–314 (1988).
- Démoulin, P. *et al.* Quasi-separatrix layers in solar flares. I. Method. *Astron. Astrophys.* **308**, 643–655 (1996).
- Li, Y., Qiu, J., Longcope, D. W., Ding, M. D. & Yang, K. Observations of an X-shaped ribbon flare in the Sun and its three-dimensional magnetic reconnection. *Astrophys. J.* **823**, L13 (2016).
- Mandrini, C. H. *et al.* Evidence of magnetic reconnection from H-alpha, soft X-ray and photospheric magnetic field observations. *Sol. Phys.* **174**, 229–240 (1997).
- Bagalá, L. G., Mandrini, C. H., Rovira, M. G. & Démoulin, P. Magnetic reconnection: a common origin for flares and AR interconnecting arcs. *Astron. Astrophys.* **363**, 779–788 (2000).
- Romano, P., Falco, M., Guglielmino, S. L. & Mirabito, M. Observation of a 3D magnetic null point. *Astrophys. J.* **837**, 173 (2017).
- Mandrini, C. H., Schmieder, B., Démoulin, P., Guo, Y. & Cristiani, G. D. Topological analysis of emerging bipole clusters producing violent solar events. *Sol. Phys.* **289**, 2041–2071 (2014).
- Dalmasse, K., Chandra, R., Schmieder, B. & Aulanier, G. Can we explain atypical solar flares? *Astron. Astrophys.* **574**, A37 (2015).
- Jiang, C., Feng, X., Wu, S. T. & Hu, Q. Study of the three-dimensional coronal magnetic field of active region 11117 around the time of a confined flare using a data driven CESE-MHD model. *Astrophys. J.* **759**, 85 (2012).
- Amari, T. & Luciani, J. F. Helicity redistribution during relaxation of astrophysical plasmas. *Phys. Rev. Lett.* **84**, 1196–1199 (2000).
- Török, T. & Kliem, B. Confined and ejective eruptions of kink-unstable flux ropes. *Astrophys. J.* **630**, L97–L100 (2005).
- Kliem, B., Titov, V. S. & Török, T. Formation of current sheets and sigmoidal structure by the kink instability of a magnetic loop. *Astron. Astrophys.* **413**, L23–L26 (2004).

21. Pinto, R., Gordovskyy, M., Browning, P. K. & Vilmer, N. Thermal and non-thermal emission from reconnecting twisted coronal loops. *Astron. Astrophys.* **585**, A159 (2016).
22. Hassanin, A. & Kliem, B. Helical kink instability in a confined eruption. *Astrophys. J.* **832**, 106 (2016).
23. Kumar, P., Yurchyshyn, V., Wang, H. & Cho, K. S. Formation and eruption of a small flux rope in the chromosphere observed by NST, IRIS and SDO. *Astrophys. J.* **809**, 83 (2015).
24. Chen, H. *et al.* Confined flare in solar active region 12192 from October 18 to 29. *Astrophys. J.* **808**, L24 (2015).
25. Myers, C. E. *et al.* A dynamic magnetic tension force as the cause of failed solar eruptions. *Nature* **528**, 526–529 (2015).
26. Lee, K.-S., Imada, S., Watanabe, K., Bamba, Y. & Brooks, D. H. *IRIS*, *Hinode*, *SDO*, and *RHESSI* observations of a white light flare produced directly by non-thermal electrons. *Astrophys. J.* **836**, 150 (2017).
27. Amari, T. *et al.* in *ASP Conference Series* (eds Pogorelov, N. V., Font, J. A., Audit, E. & Zank, G. P.) Vol. 459, 189 (Astronomical Society of the Pacific, 2012).
28. DeRosa, M. L. *et al.* The influence of spatial resolution on nonlinear force-free modeling. *Astrophys. J.* **811**, 107 (2015).
29. Amari, T., Canou, A. & Aly, J.-J. Characterizing and predicting the magnetic environment leading to solar eruptions. *Nature* **514**, 465–469 (2014).
30. Amari, T., Luciani, J. F. & Joly, P. A preconditioned semi implicit scheme for magnetohydrodynamics equations. *SIAM J. Sci. Comput.* **21**, 970–986 (1999).

**Acknowledgements** We were granted access to the High Performance Computing resources of the Centre Informatique National de l'Enseignement Supérieur (CINES) and of the Institut du Développement et de Ressources en Informatique (IDRIS) under allocation 2016-16050438 made by Grand Equipement National de Calcul Intensif (GENCI) and also to the mesocentre Phymat of the Centre National de la Recherche Scientifique/École Polytechnique. We acknowledge support of the Centre National d'Etudes Spatiales (CNES) and of the Direction Générale de l'Armement (DGA). T.A. thanks R. Huart for discussions. The Solar Dynamics Observatory (SDO) data are courtesy of the National Aeronautics and Space Administration (NASA), and the SDO/HMI and AIA science teams.

**Author Contributions** T.A. and A.C. planned and performed the various calculations and analysis discussed with J.-J.A. T.A. and F.A. worked on the mesh adaptation strategy while F.D. worked with T.A. on MESHMH. The manuscript was written by T.A. and J.-J.A. with feedback from A.C.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.A. ([tahar.amari@polytechnique.edu](mailto:tahar.amari@polytechnique.edu)).

## METHODS

**Input data.** The spherical magnetic data (Fig. 1) we use are composite, obtained from several patches of the vector magnetic field using the following technical procedure. To create the dataset, we use three different SDO/HMI products: the synoptic map of Carrington Rotation 2156 (a method of numbering the solar rotations as seen from Earth since 1853), the line-of-sight full-disk magnetogram in charged coupled device (CCD) coordinates, and the vector magnetograms from the SHARP data series in CCD coordinates. To avoid the impreciseness of the data obtained near the solar limb, we use only the vector magnetograms of the regions nearest to the solar active region AR12192. These vector magnetograms are transformed to spherical coordinates<sup>31</sup> and then embedded in the line-of-sight full-disk magnetogram at their respective locations in the CCD coordinates. The resulting data are then remapped from the CCD coordinates to the heliographic coordinates<sup>32</sup>, keeping only a spherical wedge 125° in longitude and 70° in latitude. Independently, the synoptic map is interpolated to a uniform latitude–longitude mesh. The polar field is fitted with a geometrical specification<sup>33</sup> to take care of the inaccuracy of the magnetic field measurements made near the solar poles. The final step is to embed this composite vector full-disk magnetogram into the synoptic map. To avoid discontinuities at the interfaces between the synoptic map and the full-disk map as well as between the fitted polar field and the unmodified synoptic field, the resulting radial component of the field,  $B_r$ , is subjected to a diffusion process. The non-uniform diffusion coefficient is taken to be larger around these interfaces than in the rest of the composite map, and it is set to zero in the full-disk area to keep  $B_r$  unchanged therein. As is required in any case<sup>28</sup>, a mesh adaptation algorithm is used to create high-resolution mesh in the region where it is needed. This is shown in Fig. 1d for the reconstruction of 21:00 UT. We thus finally obtained a large composite map containing a lot of information that needs to be handled correctly by the model to capture the physics involved.

**The pre-eruptive multi-scale environment model.** Our static calculations assume that the coronal magnetic field  $\mathbf{B}$  is force-free, that is, we ignore the effects of the plasma and gravity forces. This is a common approximation (also used, for example, in ref. 8) that is well justified by the low value of the plasma beta ( $\beta \approx 7 \times 10^{-4}$ ) in the lower solar corona<sup>34</sup>, where the eruption is occurring. This approximation may be considered valid not only inside a single active region, but also in a larger part of the low corona, including several interconnected active regions.

In our (static) model, the photosphere is represented by a sphere  $S$  of radius 1 and the corona by the domain  $\Omega$ , which is contained between  $S$  and an outer sphere  $S_2$  of radius  $r = 2.57$ .  $\Omega$  is assumed to be filled up with low- $\beta$ , slightly resistive and viscous plasma embedded in the magnetic field  $\mathbf{B}$ .  $\mathbf{B}$  obeys the standard force-free equations<sup>7</sup>

$$\nabla \times \mathbf{B} = \alpha \mathbf{B} \quad (1)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (2)$$

$$\mathbf{B} \cdot \nabla \alpha = 0 \quad (3)$$

where  $\alpha = \alpha(r)$  is some scalar function that is constant along field lines according to equation (3). Beyond  $S_2$ , the physics should be dominated by the solar wind, which drives the magnetic field almost radially in the heliosphere.  $S_2$  is then called the source surface.

Equations (1) to (3) are solved in  $\Omega$  using an algorithm exploiting its mixed elliptic–hyperbolic structure. This algorithm has been well described in both Cartesian and spherical<sup>35</sup> geometry. It requires as boundary conditions the normal component  $B_r$  of  $\mathbf{B}$  on  $S$ , the vanishing on  $S_2$  of the tangential component of  $\mathbf{B}$ , which thus continuously matches the outer radial field, and an assignment of the value of  $\alpha$  to each magnetic line connecting the parts  $S^+$  and  $S^-$  of  $S$ , where  $B_r > 0$  and  $B_r < 0$ , respectively. Here, we set  $\alpha$  to be equal on any line to a weighted combination of the values  $\alpha_{\text{obs}}^+$  and  $\alpha_{\text{obs}}^-$  given by the observations at its two footpoints on  $S$ , with the weights (summing to 1) reflecting our relative confidence in the precision of these data. The actual computation of  $\mathbf{B}$  is done by imposing such a boundary condition at each hyperbolic step of the iterative scheme<sup>36,37</sup>.

To take advantage of the richness of these high-resolution data, our multi-scale method solves equations (1) to (3) numerically using a scheme defined on an unstructured tetrahedral mesh that can be adapted along the iterative algorithm lying at the heart of our MESHMHD numerical code. To begin with, we look for the best triangular mesh adapted to the data given on the photosphere. This is done by starting with a crude triangular mesh and using a fixed-point algorithm. Adaptation followed by data interpolation (for  $\alpha$  and  $B_r$ ) on the new mesh thus leads to an optimal mesh associated with the data, with high resolution where data information requires small triangles and lower resolution elsewhere. This complex procedure is needed to obtain the high-quality mesh shown in Fig. 1, which is then used to build an initial tetrahedral mesh inside  $\Omega$ . Our iterative algorithm<sup>27</sup> is next

started by using this initial mesh. After each iteration, quantities called sensors, such as the magnetic field  $\mathbf{B}$  and  $\alpha$ , are used to adapt the mesh where small-scale structures form, such as electric current concentrations and the TFR (Fig. 2). This is continued using a fixed-point method<sup>38</sup> until the couple mesh/solution converges to a small interpolation error.

Using this model, we obtain an equilibrium with very good standard diagnostics. The evolution and asymptotic convergence along the iterative scheme<sup>39</sup> have been computed for two of these diagnostics (Extended Data Fig. 5):  $(\mathbf{J}, \mathbf{B})$  and  $\langle |f_i| \rangle$ .  $(\mathbf{J}, \mathbf{B})$  is the angle between the electric current density  $\mathbf{J}$  and  $\mathbf{B}$ , which measures the extent to which  $\mathbf{B}$  is force-free, and the standard parameter  $\langle |f_i| \rangle$  provides an estimate of the divergence of  $\mathbf{B}$ . When convergence is obtained,  $(\mathbf{J}, \mathbf{B})$  is smaller than 2°, while  $\langle |f_i| \rangle \approx 10^{-17}$ . This means that the constraint  $\nabla \cdot \mathbf{B} = 0$  is satisfied to a high degree of precision ( $|\nabla \cdot \mathbf{B}|$  is at most equal to about  $10^{-12} \text{ G Mm}^{-1}$ ). This last result is a direct consequence of the structure of our numerical algorithm, which rests on a discretization of the equations that makes the magnetic field exactly divergence-free. We work in the functional spaces for element representation that are divergence-free on each tetrahedron, which ensures that  $\mathbf{B}$  is in the kernel of the divergence operator.  $\mathbf{B}$  is computed as the curl of a vector potential  $\mathbf{A}$ , with  $\mathbf{B}$  and  $\mathbf{A}$  defined on each tetrahedron, while the divergence operator is defined by duality.

To have such low values for these diagnostics is a necessary condition for the electric currents to organize themselves into a well formed TFR with high twist (Fig. 2) during the reconstruction process performed by our model, which increases the resolution of the mesh where the current and the magnetic field change over smaller scales. Moreover, comparisons with coronal and ribbon data provided by emission observations made by AIA (onboard the NASA/SDO mission) show interesting correlations (see Extended Data Figs 6 and 7).

**Auxiliary force-free field  $\mathbf{B}'$ .** Electric currents flowing outside the TFR (denoted  $T$ ) of the reconstructed force-free field  $\mathbf{B}$  are far from being negligible in the active region we consider. This appears quite clearly in Extended Data Fig. 2, where we can see electric currents of both signs associated with the TFRs that fill the volume around or above  $T$ . But unfortunately their effects have not yet been taken into account in the theoretical force-free models that are generally used for establishing the stability criteria of a TFR. To quantify the importance of the electric currents, we introduced a new auxiliary force-free magnetic field,  $\mathbf{B}'$ , constructed in such a way that all its electric currents are fully confined inside a TFR (denoted  $T'$ ) that differs as little as possible from  $T$ .

Let us denote as  $s^{+/-}$  the intersection of  $T$  with the part  $S^{+/-}$  of positive/negative polarity of the solar surface. To compute  $\mathbf{B}'$ , we first solve the same Grad–Rubin boundary value problem as for  $\mathbf{B}$ , but with a change in the boundary conditions: on  $S^+$ , we require  $\alpha$  to vanish outside  $s^+$  instead of assuming the observed value. In this way, we obtain a field whose currents are concentrated in a tube having the same intersection with  $S^+$  as  $T$ , but an intersection with  $S^-$  different from  $s^-$ . Therefore, we set up an iterative procedure to transform that field into another field  $\mathbf{B}'$  that has a TFR  $T'$  as close as possible to  $T$ . Each iterative step consists in solving a Grad–Rubin problem in which an appropriate value of  $\alpha$  is imposed either on  $S^+$  or  $S^-$ .

**Magnetohydrodynamic evolution.** Once several reconstructions have been made in the four hours before the eruption, we have several different equilibrium states that may be a priori expected to be stable, because otherwise the reconstruction scheme would not have converged. We therefore undertake a new step in which, with the help of a magnetohydrodynamic code, we determine how each of these equilibrium states evolves under the action of several processes that are known to occur on the solar photosphere: flux cancellation, flow convergence and turbulent diffusion. These processes are mimicked by properly chosen boundary conditions, and the effects of each of them are considered in turn. We do not perform a standard stability analysis<sup>8</sup>, in which one would study how an initial equilibrium state evolves either under the effects of some arbitrary superposed perturbations, or simply from the interpolation errors left when bringing data from one code to the other. Rather, we impose physically motivated nonlinear evolutions on our magnetic structure, which stay close to the actual evolution of the active region because the computations are run for only a short time.

The tool we use is our efficient magnetohydrodynamics code METEOSOL<sup>30</sup>, which has already been extensively employed and has produced many results<sup>18,40,41</sup>, in particular in theoretical studies of the effects on a given magnetic structure of the surface processes mentioned above. Since the big flare that occurred around 21:00 UT was not ejective, the calculations are performed on a high-resolution METEOSOL mesh set in a large Cartesian box (such a box has been proved by previous simulations of coronal mass ejections not to exert any spurious confining effect on the magnetic field, thus ensuring that the resulting eruption is physically confined and not artificially confined by the box). The mesh used by METEOSOL is a staggered Cartesian non-uniform one. The spatial discretization of the operators is defined in such a way that magnetic helicity

and topology are conserved in the weakly resistive limit and that the constraint  $\nabla \cdot \mathbf{B} = 0$  is satisfied, to round off errors. METEOSOL uses an efficient time-advance semi-implicit scheme. Some modifications have been introduced since the original version of METEOSOL<sup>30</sup>. In particular, the convective terms are treated by a third-order non-oscillatory upwind scheme. The model uses several different assumptions and options, among which a classical zero- $\beta$  approach, with an imposed profile of density that is either constant or decreasing with altitude. We have also tried an initial plasma hydrostatic equilibrium, assuming the adiabatic approximation in which the heating terms are neglected. This led to only minor changes in the results. For the viscosity and the resistivity, we choose values between  $10^{-2}$  and  $10^{-3}$  (in code units) and in the range  $10^{-4}$  down to  $10^{-5}$ , respectively. The results were found to be independent of the values selected for these parameters.

Applying the method above, we find that the evolution of the pre-eruptive configuration of 21:00 UT is driven by any of the three surface mechanisms previously listed, which leads after a short time to the development of the kink instability of the TFR, which otherwise stays confined (Fig. 4). No instability is found to develop during a short period of time when the same procedure is applied to earlier static configurations, which we observed to evolve slightly and inconsequentially.

A further result concerns the effect of applying flux cancellation for a longer time and on a larger scale including the magnetic cage. In that case, the resulting decrease in the confinement leads to a major disruption of the structure and to an ejection of the TFR, that is, a coronal mass ejection is produced as in our previous work<sup>29</sup> (Extended Data Fig. 9). This result provides one more illustration of the confining power of the cage and then of the effect of the environment on the nature of the event (a confined flare versus eruptive flare plus coronal mass ejection) that can be produced.

During the development of the instability of the 21:00 UT (hour H) pre-eruptive configuration, magnetic helicity is conserved, keeping a value of about  $-300 \times 10^{42} \text{ Mx}^2$  ( $1 \text{ maxwell} = 1 \text{ G cm}^2$ ; Fig. 4d). However, the configuration does not relax between the pre-eruptive and post-eruptive phases to a linear force-free field having the same helicity; that is, the Taylor conjecture<sup>42</sup> is not valid here. Instead, nonlinearities are present in the predicted post-eruptive state, which is otherwise found to be close to the state obtained by applying

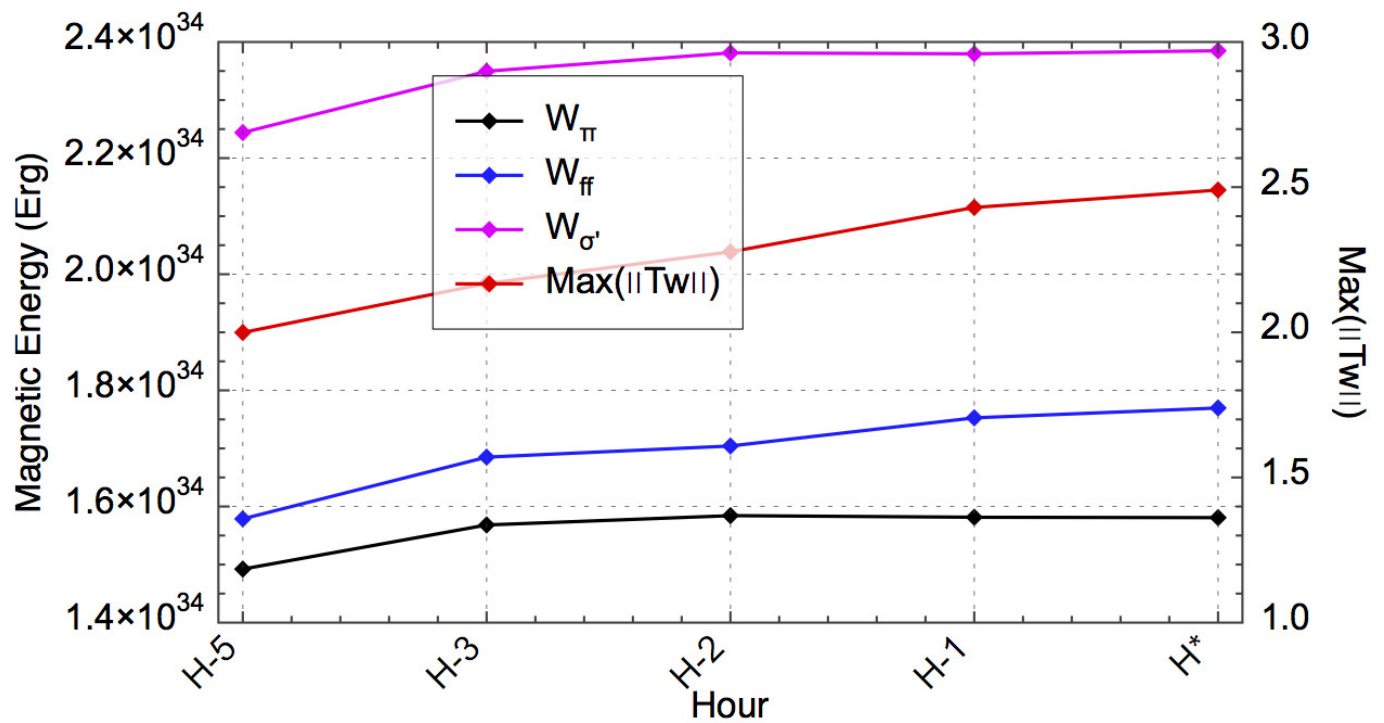
our static model to the HMI observations made on 25 October 2014 (Extended Data Fig. 10).

**Code availability.** We have decided not to make our code available. Several parts of the code we used are embedded in a complicated way, and are the private property of the authors' institutions. The reader, however, should be able to reproduce our results by using the information provided in Methods.

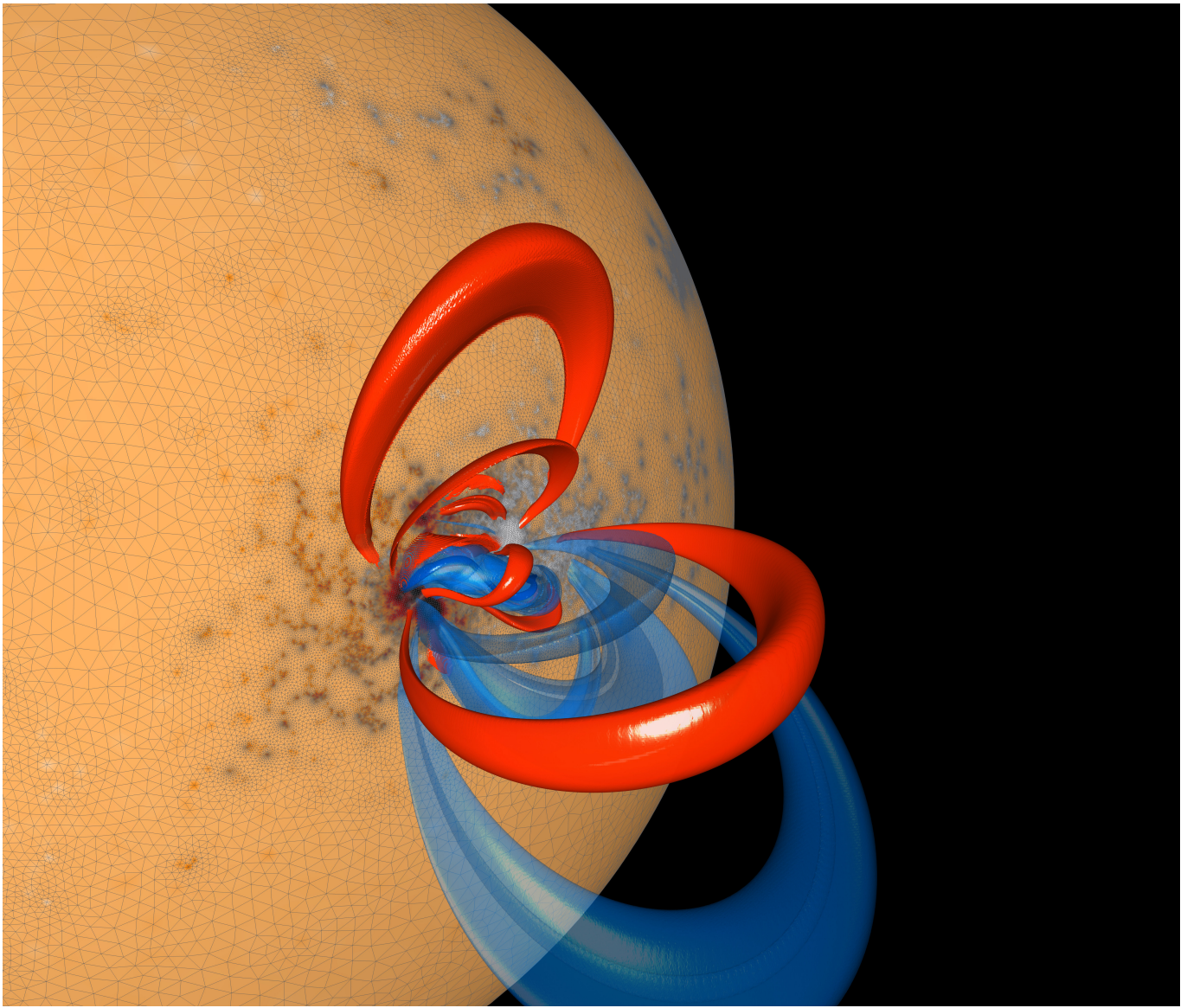
**Data availability.** The data files that support our analysis will be made available upon reasonable request.

31. Allen Gary, G. & Hagyard, M. J. Transformation of vector magnetograms and the problem associated with the effects of perspective and the azimuthal ambiguity. *Sol. Phys.* **126**, 21–36 (1990).
32. Thompson, W. T. Coordinate systems for solar image data. *Astron. Astrophys.* **449**, 791–803 (2006).
33. Titov, V. S., Mikic, Z., Török, T., Linker, J. A. & Panasenco, O. 2010 August 1–2 sympathetic eruptions. I. Magnetic topology of the source-surface background field. *Astrophys. J.* **759**, 70 (2012).
34. Priest, E. R. *Magnetohydrodynamics of the Sun* (Cambridge Univ. Press, 2014).
35. Amari, T., Aly, J.-J., Canou, A. & Mikic, Z. Reconstruction of the solar coronal magnetic field in spherical geometry. *Astron. Astrophys.* **553**, A43 (2013).
36. Amari, T. & Aly, J.-J. Observational constraints on well-posed reconstruction methods and the optimization-Grad-Rubin method. *Astron. Astrophys.* **522**, A52 (2010).
37. Wheatland, M. S. & Régnier, S. A Self-consistent nonlinear force-free solution for a solar active region magnetic field. *Astrophys. J.* **700**, L88–L91 (2009).
38. Alauzet, F., Frey, P. J., George, P. L. & Mohammadi, B. 3D transient fixed point mesh adaptation for time-dependent problems: application to CFD simulations. *J. Comput. Phys.* **222**, 592–623 (2007).
39. Guo, Y., Xia, C., Keppens, R. & Valori, G. Magneto-frictional modeling of coronal nonlinear force-free fields. I. Testing with analytic solutions. *Astrophys. J.* **828**, 82 (2016).
40. Amari, T., Luciani, J.-F., Aly, J.-J., Mikic, Z. & Linker, J. Coronal mass ejection: initiation, helicity and flux ropes. II. *Turbulent diffusion driven evolution*. *Astrophys. J.* **595**, 1231–1250 (2003).
41. Amari, T., Aly, J.-J., Luciani, J.-F., Mikic, Z. & Linker, J. Coronal mass ejection initiation by converging photospheric flows: toward a realistic model. *Astrophys. J.* **742**, L27 (2011).
42. Taylor, J. B. Relaxation of toroidal plasma and generation of reverse magnetic fields. *Phys. Rev. Lett.* **33**, 1139–1141 (1974).

## Magnetic Energy and Twist

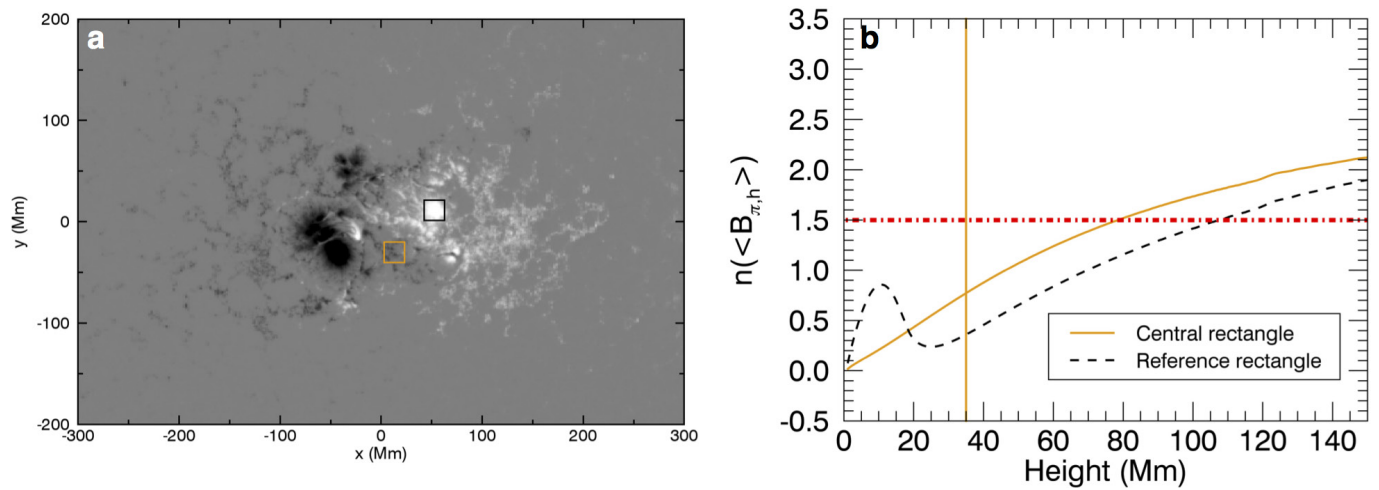


**Extended Data Figure 1 | Evolution of magnetic energies and twist.** Evolution during the four hours preceding the eruption of the actual magnetic field energy (blue), the potential field energy (black), and the semi-open field energy (purple), expressed in physical units. The evolution of the twist is also plotted (red).



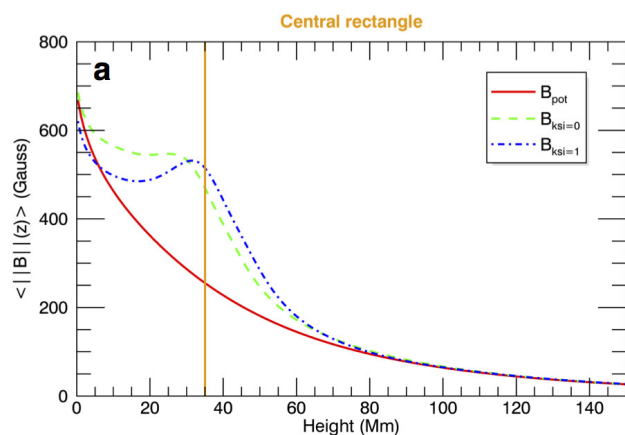
**Extended Data Figure 2 | Electric current structure of the magnetic environment.** Set of selected flux ropes, including the central highly twisted flux rope and some ropes having a non-negligible twist of more

than 0.5. These ropes are located around and above the major TFR and reveal a complex non-potential environment. Those in blue (red) are associated with a negative (positive) value of  $\alpha$ .

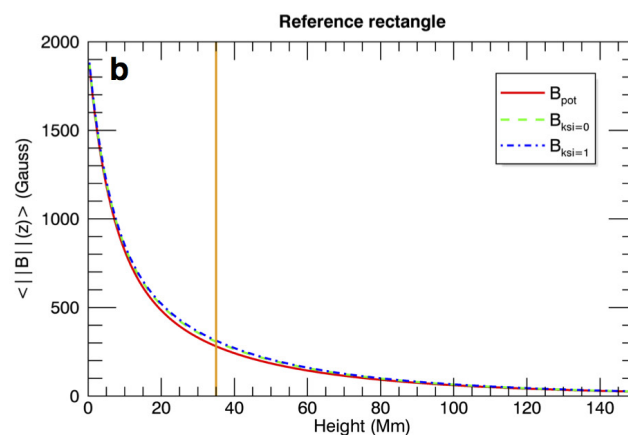


**Extended Data Figure 3 | Index of torus instability.** **a**, Magnetogram at 21:00 UT with rectangles indicating the sample area in which the index  $n$  is computed. The yellow rectangle is located just below the TFR, while the black rectangle, used as a reference, is chosen outside it. **b**, Variation with altitude of the torus index computed above the sample areas shown in **a**

(with the same colour coding) by using the horizontal component<sup>19</sup> of the mean potential (current-free) magnetic field,  $\langle B_{\pi,h} \rangle$ . The horizontal line indicates the critical value of the index often used for the torus instability, while the vertical one indicates the height of the TFR axis.

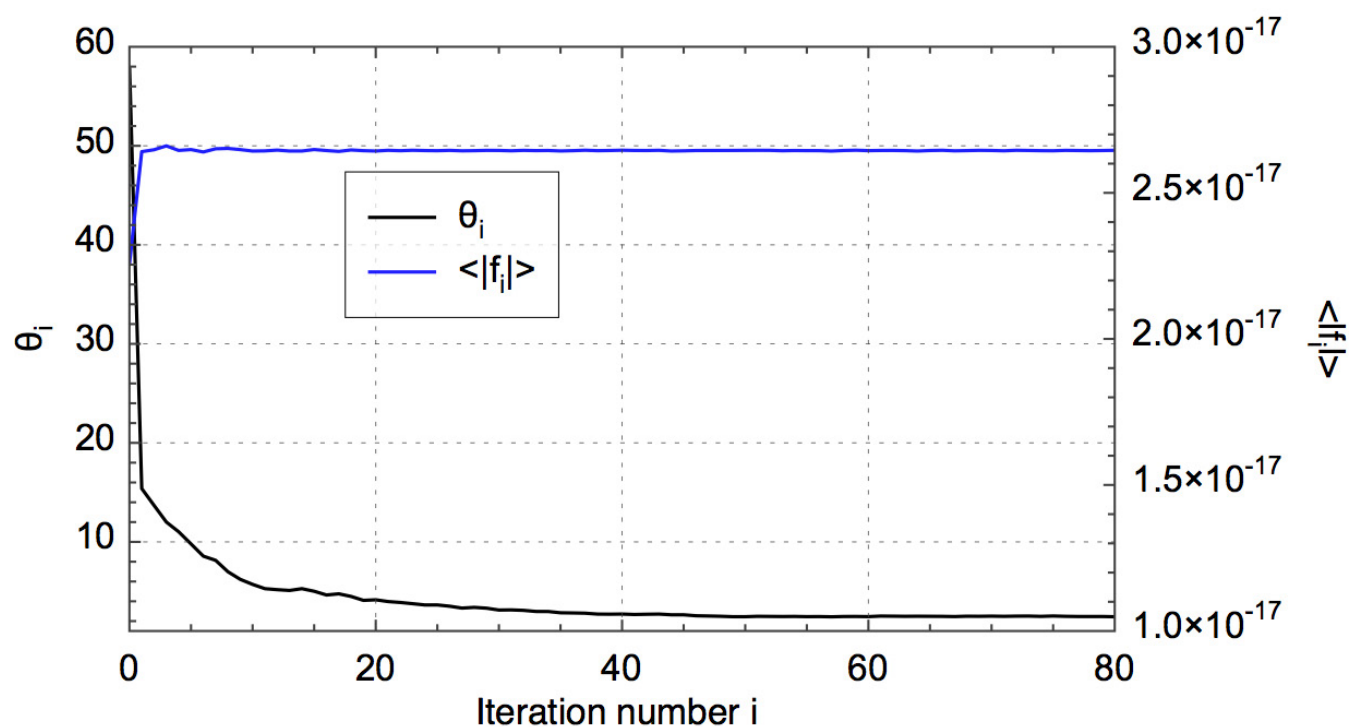


**Extended Data Figure 4 | Role of volumetric current.** Variation of the magnetic intensity mean value above the rectangles shown in Extended Data Fig. 3a for the current-free solution (red curve), the force-free solution  $B'$ , computed by removing the photospheric electric current exterior to the base of the TFR (green curve), and the full force-free solution ( $B$ ), computed by taking into account the total electric current



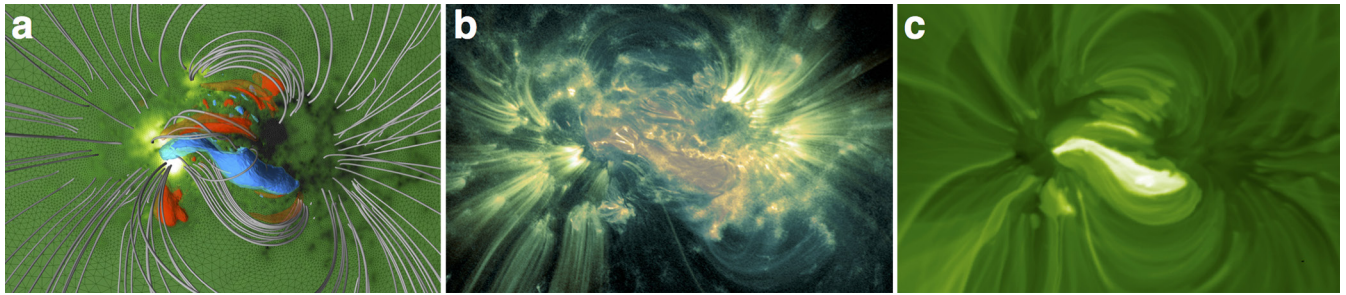
(blue curve). The number ksi measures the degree of removal of those external currents, with  $\text{ksi} = 0$  ( $\text{ksi} = 1$ ) indicating total (no) removal. As in Extended Data Fig. 3, the vertical yellow line indicates the height of the TFR axis. **a**, The central rectangle; **b**, the reference rectangle outside the TFR area.

## Convergence Diagnostics



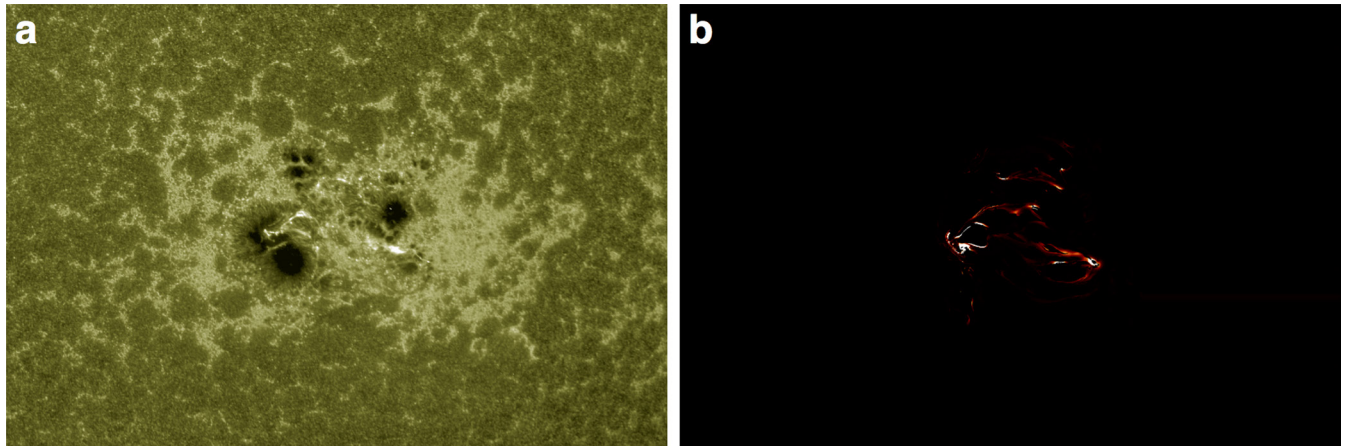
**Extended Data Figure 5 | Quality of the reconstruction method.**  
Evolution of the angle  $\theta_i$  between the electric current vector and the magnetic field (a good measure of how force-free the solution is) during

the iterations of the algorithm, as well as that of the diagnostic standard parameter  $\langle |f_i| \rangle$ , measuring the divergence of the solution.  $i$  is the iteration number.



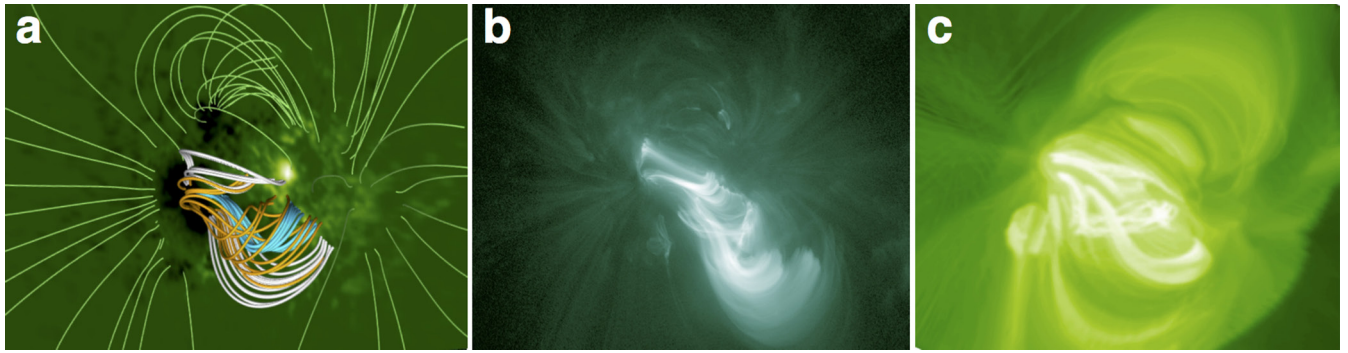
**Extended Data Figure 6 | Comparison with AIA data.** **a**, Selected field lines and set of different isosurfaces of the force-free scalar function  $\alpha$  (red for positive values and blue for negative values) for the reconstructed pre-eruptive magnetic configuration of 24 October 2014 at 21:00 UT.

**b**, Corresponding composite image from the AIA-131 Å and AIA-171 Å wavelength data. **c**, Synthetic emissivity computed by using the magnetic field and the electric current density of the reconstructed pre-eruptive magnetic configuration of 24 October 2014 at 21:00 UT.

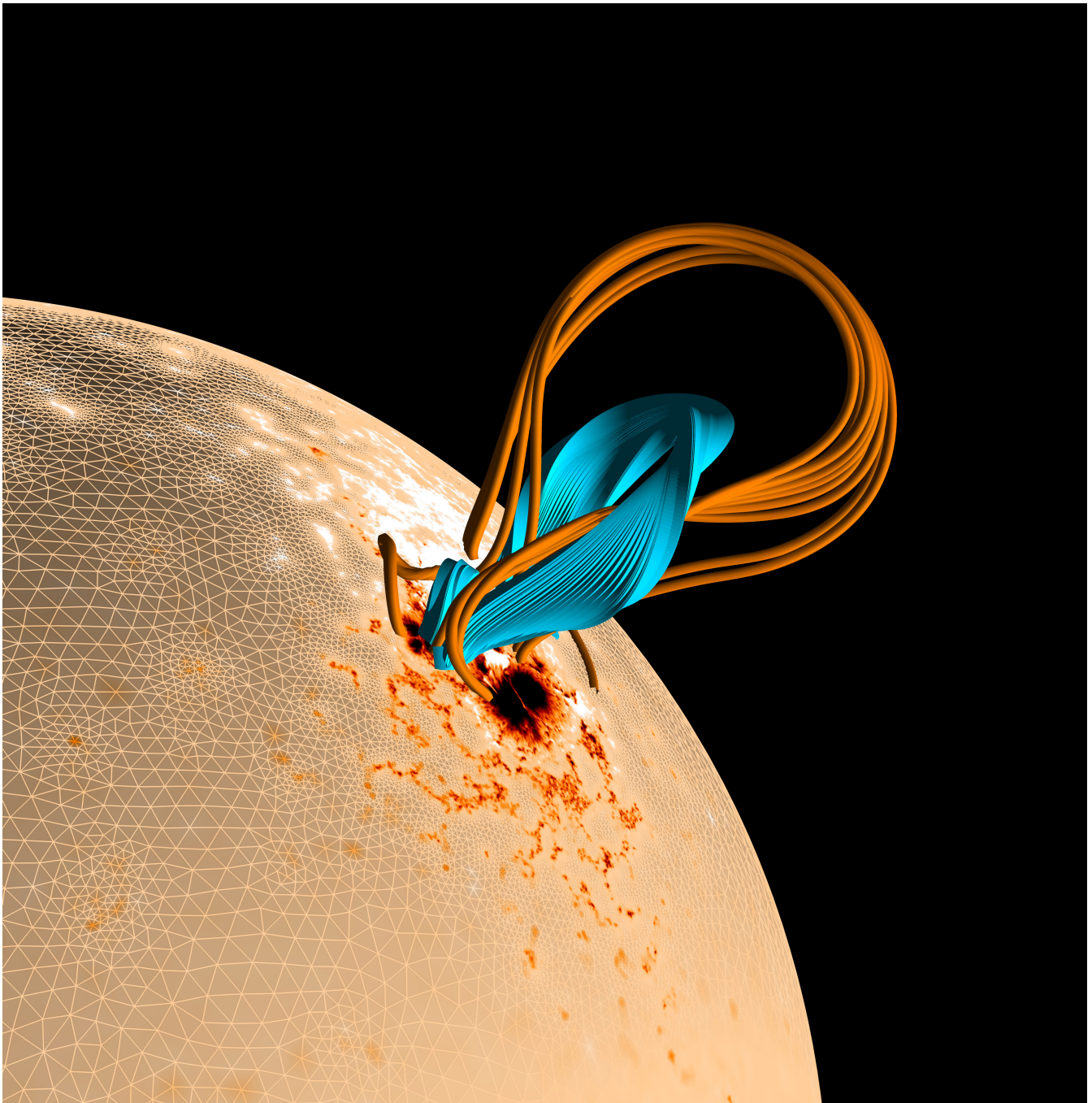


**Extended Data Figure 7 | Comparison with AIA-1,600 Å wavelength.** **a**, AIA emission at 1,600 Å. **b**, Plot of the vertically integrated dissipation  $J^2$  (where  $J$  is the norm of the electric current density) above the regions with high values of the squashing factor<sup>33</sup>, as for the reconstructed

pre-eruptive magnetic configuration of 24 October 2014 at 21:00 UT. This plot highlights the strong electric current regions, in which reconnection is expected to occur.

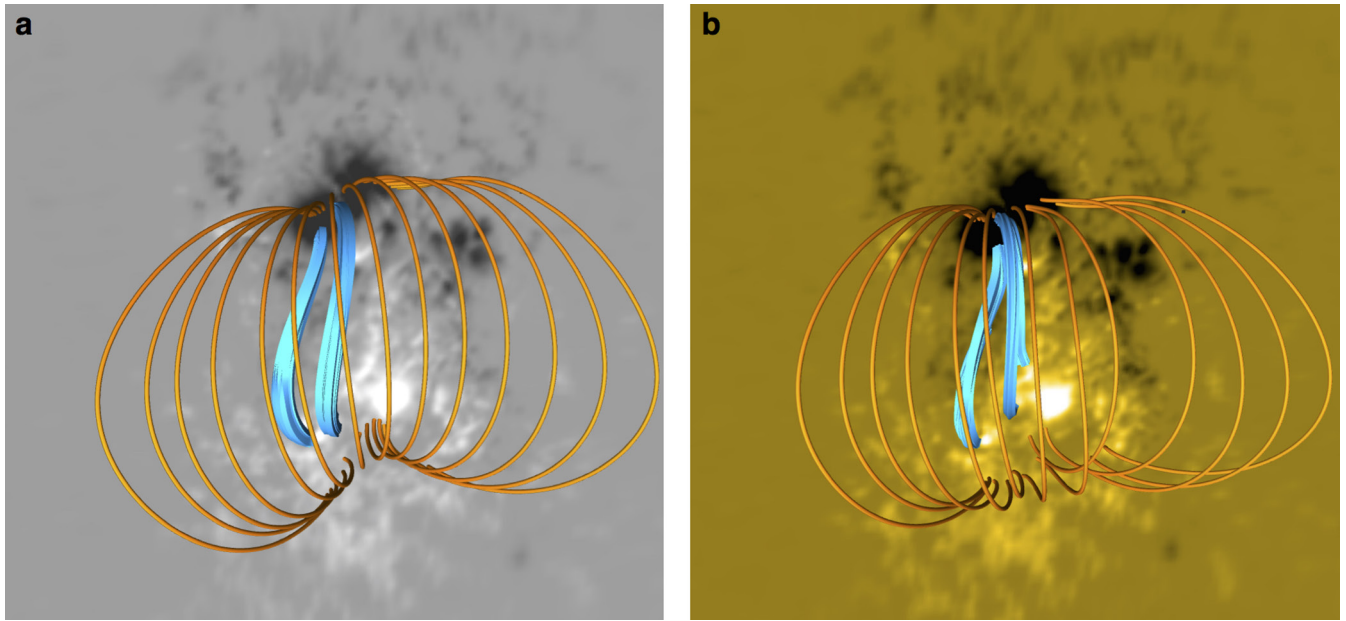


**Extended Data Figure 8 | Extreme-ultraviolet emission and magnetic structure.** **a**, Selected field lines of the evolving magnetic configuration during the flare. **b**, AIA emission at 94 Å on 24 October 2014 at 22:00 UT. **c**, Synthetic emissivity computed by using the magnetic field and the electric current density of the evolving magnetic configuration during the flare.



**Extended Data Figure 9 | Major eruption and role of the magnetic environment.** Selected field lines of the configuration that evolved into a major coronal mass ejection from the pre-eruptive configuration of 24

October 2014 at 21:00 UT when flux cancellation was applied on a larger scale, including the magnetic cage, whose confinement effect has thus been weakened.



**Extended Data Figure 10 | Post-eruptive state.** Comparison of the post-eruptive states obtained from the simulation after the full relaxation of the evolving unstable state (a) and using HMI vector magnetic data from 25 October 2014 (b).

# Isomer depletion as experimental evidence of nuclear excitation by electron capture

C. J. Chiara<sup>1</sup>, J. J. Carroll<sup>2</sup>, M. P. Carpenter<sup>3</sup>, J. P. Greene<sup>3</sup>, D. J. Hartley<sup>4</sup>, R. V. F. Janssens<sup>3†</sup>, G. J. Lane<sup>5</sup>, J. C. Marsh<sup>1‡</sup>, D. A. Matters<sup>6</sup>, M. Polasik<sup>7</sup>, J. Rządkiwicz<sup>8</sup>, D. Seweryniak<sup>3</sup>, S. Zhu<sup>3</sup>, S. Bottoni<sup>3‡</sup>, A. B. Hayes<sup>9</sup> & S. A. Karamian<sup>10‡</sup>

The atomic nucleus and its electrons are often thought of as independent systems that are held together in the atom by their mutual attraction. Their interaction, however, leads to other important effects, such as providing an additional decay mode for excited nuclear states, whereby the nucleus releases energy by ejecting an atomic electron instead of by emitting a  $\gamma$ -ray. This ‘internal conversion’ has been known for about a hundred years and can be used to study nuclei and their interaction with their electrons<sup>1–3</sup>. In the inverse process—nuclear excitation by electron capture (NEEC)—a free electron is captured into an atomic vacancy and can excite the nucleus to a higher-energy state, provided that the kinetic energy of the free electron plus the magnitude of its binding energy once captured matches the nuclear energy difference between the two states. NEEC was predicted<sup>4</sup> in 1976 and has not hitherto been observed<sup>5,6</sup>. Here we report evidence of NEEC in molybdenum-93 and determine the probability and cross-section for the process in a beam-based experimental scenario. Our results provide a standard for the assessment of theoretical models relevant to NEEC, which predict cross-sections that span many orders of magnitude. The greatest practical effect of the NEEC process may be on the survival of nuclei in stellar environments<sup>7</sup>, in which it could excite isomers (that is, long-lived nuclear states) to shorter-lived states. Such excitations may reduce the abundance of the isotope after its production. This is an example of ‘isomer depletion’, which has been investigated previously through other reactions<sup>8–12</sup>, but is used here to obtain evidence for NEEC.

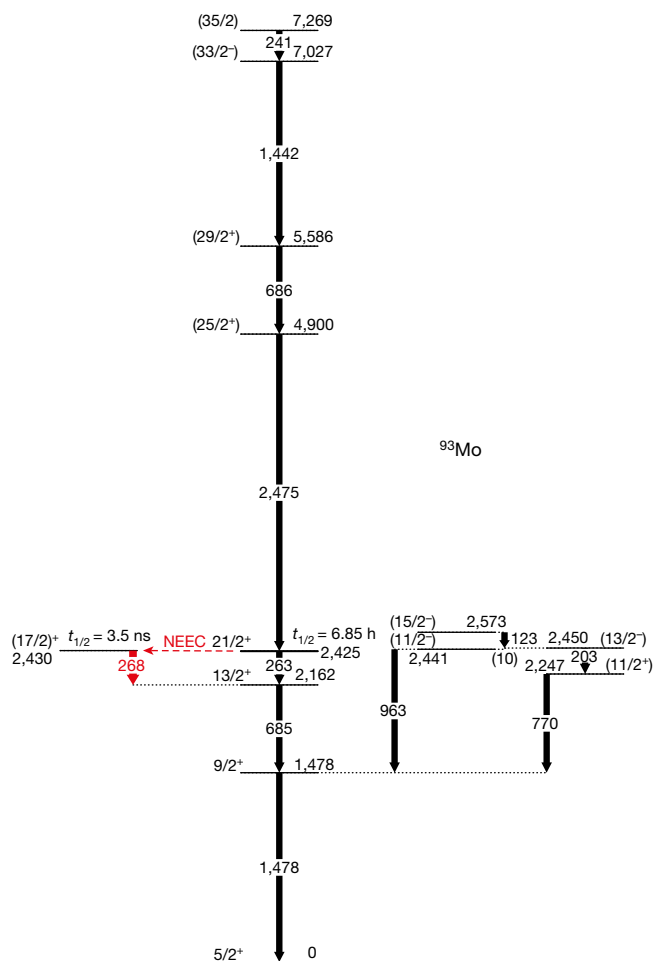
We searched for evidence of the NEEC process in <sup>93</sup>Mo following an approach adapted from a technique proposed in ref. 13. Previous theoretical work also considered <sup>93</sup>Mo as a NEEC candidate<sup>6</sup>. This nucleus has a 21/2<sup>+</sup> isomer at 2,425 keV with a half-life of 6.85 h and a (17/2)<sup>+</sup> candidate intermediate state that lies 4.85(9) keV higher at 2,430 keV (ref. 14), as illustrated in Fig. 1. (Uncertainties are quoted at the 1 $\sigma$  or 68% confidence level, unless otherwise noted.) In our approach, <sup>93</sup>Mo is produced in its metastable state (<sup>93m</sup>Mo) through nuclear reactions. Choosing heavy projectiles and light target nuclei results in recoiling <sup>93m</sup>Mo reaction products (‘recoils’) that move at high velocities  $v$  (initially more than 10% of the speed of light,  $c$ ) in approximately the same direction as the beam. As the fast-moving recoils pass through the target medium, electrons are stripped off, leaving the <sup>93m</sup>Mo ions with a high average charge<sup>15,16</sup> between about +32 and +36. Subsequent collisions with target atoms reduce the energy of the recoiling ions while simultaneously providing electrons that can be captured back into the vacated atomic orbitals. At the right combination of the charge state

of the <sup>93m</sup>Mo ion and the effective electron kinetic energy, as seen from the reference frame of the recoiling ion, capture can occur, releasing enough energy to match that needed ( $\Delta E = 4.85$  keV) to excite the nucleus from the isomer to the intermediate state—that is, NEEC occurs. As described in detail by recent theoretical calculations<sup>16</sup>, these NEEC conditions could be fulfilled in practice by capture of an electron at the necessary relative kinetic energy, to within the line width of the capturing atomic vacancy (the NEEC resonance width). The intermediate state in <sup>93</sup>Mo that would be populated through NEEC is known to decay (with a half-life of  $t_{1/2} = 3.5$  ns) to the ground state through a characteristic sequence of  $\gamma$ -rays at 268 keV, 685 keV and 1,478 keV (Fig. 1). Notably, the 268-keV transition would never be seen in the natural decay of the 21/2<sup>+</sup> isomer.

The <sup>90</sup>Zr + <sup>7</sup>Li fusion–evaporation reaction was selected for our experiment, which was performed at the ATLAS facility at Argonne National Laboratory. An 840-MeV <sup>90</sup>Zr beam was provided with an average beam intensity of about  $6 \times 10^8$  ions s<sup>−1</sup>. The multi-layer target was composed of <sup>7</sup>Li supported by a natural carbon (<sup>nat</sup>C) foil, followed by a gap of about 3 mm and an additional <sup>nat</sup>C layer, backed with <sup>208</sup>Pb (see Methods for the choice of reaction and target construction). The target was positioned at the centre of the Gammasphere  $\gamma$ -ray spectrometer<sup>17</sup>, which, at the time of the experiment, comprised 92 Compton-suppressed, high-purity Ge detectors arranged in 16 rings of constant angle relative to the beam direction. A minimum of three  $\gamma$ -rays within a 2- $\mu$ s coincidence window was required for events to be recorded to disk, although a narrower, sub-microsecond coincidence constraint was imposed in the offline analysis. We operated the array with a digital data acquisition system (Digital Gammasphere<sup>18</sup>) at average event rates of 40–50 kHz. We collected data for about 62 h in this configuration.

After the fusion of the <sup>90</sup>Zr projectiles and the <sup>7</sup>Li target nuclei, and the subsequent evaporation of a proton and three neutrons from the compound nuclear system, excited states above the <sup>93m</sup>Mo isomer were populated. These excited states depopulated through decay paths that, in some cases, fed the 21/2<sup>+</sup> isomeric state, such as through the 2,475-keV, (25/2<sup>+</sup>)  $\rightarrow$  21/2<sup>+</sup> transition (parentheses denote tentative assignments). If a recoiling <sup>93</sup>Mo ion is in this isomeric state when the energy–charge resonance conditions are met, NEEC may occur, shifting some of the population of the isomer to the 3.5-ns intermediate state. An observable signature of NEEC would then be the detection of a 2,475-keV  $\gamma$ -ray, which would confirm that the nucleus had reached the <sup>93m</sup>Mo state, within the same coincidence window as one or more  $\gamma$ -rays from the decay of the intermediate level. The 2,475-keV

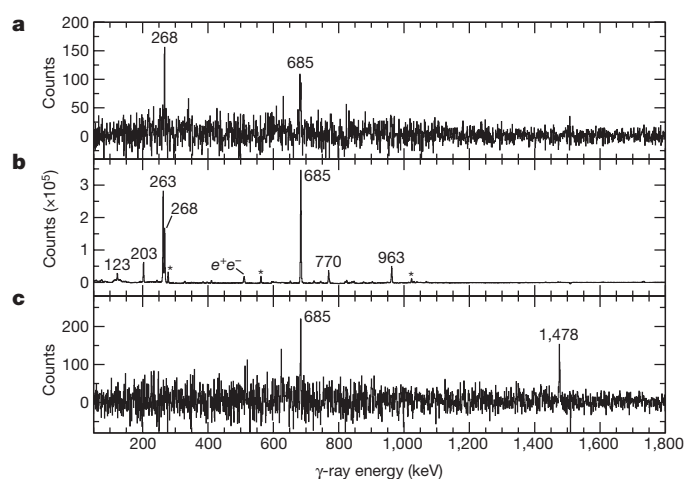
<sup>1</sup>Oak Ridge Associated Universities Fellowship Program, US Army Research Laboratory, Adelphi, Maryland 20783, USA. <sup>2</sup>US Army Research Laboratory, Adelphi, Maryland 20783, USA. <sup>3</sup>Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. <sup>4</sup>Department of Physics, US Naval Academy, Annapolis, Maryland 21402, USA. <sup>5</sup>Department of Nuclear Physics, Research School of Physics and Engineering, Australian National University, Canberra, Australian Capital Territory 0200, Australia. <sup>6</sup>Defense Threat Reduction Agency, Fort Belvoir, Virginia 22060, USA. <sup>7</sup>Faculty of Chemistry, Nicolaus Copernicus University in Toruń, 87-100 Toruń, Poland. <sup>8</sup>National Centre for Nuclear Research, 05-400 Otwock, Poland. <sup>9</sup>Institute of Optics, University of Rochester, Rochester, New York 14627, USA. <sup>10</sup>Flerov Laboratory of Nuclear Reactions, Joint Institute for Nuclear Research, Dubna 141980, Russia. †Present addresses: Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3255, USA and Triangle Universities Nuclear Laboratory, Duke University, Durham, North Carolina 27708-2308, USA (R.V.F.J.); Frontier Technology, Inc., Space and Naval Warfare Systems Command, Systems Center Pacific, 49599 Lassing Road, San Diego, California 92152, USA (J.C.M.); Dipartimento di Fisica, Università degli Studi di Milano and INFN sez. Milano, I-20133, Milano, Italy (S.B.).  
‡Deceased.



**Figure 1 | Relevant part of the  $^{93}\text{Mo}$  decay scheme.** The energy and the spin and parity quantum numbers ( $I^\pi$ , where  $I$  denotes the spin and  $\pi = +, -$  the parity) are noted for each level, with tentative assignments in parentheses. The half-lives  $t_{1/2}$  given for the two isomeric states are from ref. 14. Level and  $\gamma$ -ray energies are from this study and are given in kiloelectronvolts. The 10-keV transition between the (13/2-) and (11/2-) states was not observed in this work, but was inferred from the measured coincidence relationships. The NEEC transition from the isomer to the intermediate state is indicated by the dashed red line. The key 268-keV  $\gamma$ -ray is also indicated in red.

$\gamma$ -ray would not ordinarily be found in true coincidence with the 685- or 1,478-keV transitions because of the narrow width (2  $\mu\text{s}$ ) of the coincidence window compared to the 6.85-h half-life of the metastable state, nor would it ever be coincident with the 268-keV  $\gamma$ -ray. These coincidences would be possible, however, if some mechanism, such as NEEC, excited the nucleus from the long-lived isomer to the intermediate state.

Each recorded event, consisting of the data for three or more coincident  $\gamma$ -rays, was decomposed into all the constituent subsets of three  $\gamma$ -rays. Energy conditions, or 'gates', were imposed to select two  $\gamma$ -rays out of each set of three, and the counts in a coincidence spectrum were incremented at the energy of the third photon. Background counts, determined by gating on energy ranges near the peaks of interest, were subtracted (see Methods for the gating and background-subtraction procedures, and for the impact of the Doppler effect). A background-subtracted spectrum, produced by a double gate on the 2,475- and 1,478-keV transitions, is presented in Fig. 2a. For comparison, a background-subtracted spectrum, single-gated on just the 1,478-keV  $\gamma$ -ray (that is, the second  $\gamma$ -ray can be at any energy), is also shown in Fig. 2b. In the latter spectrum, the more intense transitions that are coincident with the 1,478-keV  $\gamma$ -ray (see Fig. 1) are clearly visible. With the additional constraint of a gate on 2,475 keV, these lines disappear

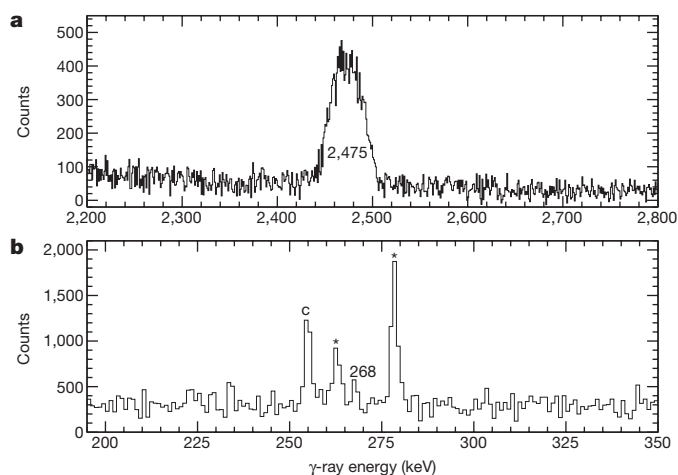


**Figure 2 | Spectra demonstrating the signature of NEEC in  $^{93}\text{Mo}$ .** No correction for the Doppler effect has been applied. **a**, Spectrum obtained with a double gate on the Doppler-shifted 2,475-keV and unshifted 1,478-keV  $\gamma$ -rays. **b**, Spectrum obtained with a single gate on the unshifted 1,478-keV line. **c**, Spectrum obtained with a double gate on the Doppler-shifted 2,475-keV and unshifted 268-keV  $\gamma$ -rays. Peaks of  $^{93}\text{Mo}$  shown in Fig. 1 are labelled with their energies in kiloelectronvolts. Additional known  $^{93}\text{Mo}$  transitions, not shown in Fig. 1, are marked with asterisks in **b**. The label  $e^+e^-$  indicates the 511-keV electron-positron annihilation peak. We note that transitions located above the isomer are too spread out in energy by the Doppler effect to be visible in these spectra.

apart from the 268- and 685-keV peaks (Fig. 2a). The counts in the key 268-keV line are about  $7\sigma$  above background, where  $\sigma$  is the standard deviation of the counts in the nearby background region (statistical error), or over  $3\sigma$  when the uncertainty associated with background subtraction is also taken into account (statistical plus systematic errors, as described in Methods). A similar gating procedure on the 2,475- and 268-keV transitions reveals peaks at 685 keV and 1,478 keV, as seen in Fig. 2c. Combined, these spectra demonstrate that the  $\gamma$ -ray sequence 2,475 keV–268 keV–685 keV–1,478 keV occurs in coincidence, which is the expected signature of isomer depletion via NEEC. We note that unlike previous examples of isomer depletion<sup>8,9,19</sup>, here the energy of the excitation to the intermediate state is much smaller than that of the isomer itself.

To determine the probability of this excitation,  $P_{\text{exc}}$ , we summed background-subtracted spectra that were double-gated on the Doppler-shifted lines at 241 keV and 1,442 keV and at 241 keV and 686 keV (see Methods for details). Two versions of this spectrum were generated, as shown in Fig. 3. One had a Doppler correction applied for  $v/c = 0.109$  recoils (Fig. 3a), so that the area of the 2,475-keV peak could be determined. The other spectrum had no Doppler correction (Fig. 3b) and was used to obtain the area of the 268-keV peak. As there is no identified decay path from the gating transitions to the 268-keV  $\gamma$ -ray, aside from those involving an excitation such as NEEC, the ratio of the areas of the 268- and 2,475-keV peaks, when corrected for their detection efficiencies, yields the probability that a  $^{93\text{m}}\text{Mo}$  ion in the long-lived isomeric state (reached via the 2,475-keV transition) subsequently emits a 268-keV  $\gamma$ -ray. An additional, small correction for internal conversion provides the excitation probability, which we determined to be  $P_{\text{exc}} = 0.010(3)$ . The corresponding cross-section, averaged over the full target thickness, is 40 b (see Methods). This is a lower limit, however, as NEEC is a resonant process; the peak cross-section may be much larger. Given that existing cross-section calculations involve very different conditions, such as channelling in a crystal<sup>20,21</sup> or in laser-induced plasmas<sup>22–24</sup>, and predict values spanning many orders of magnitude, an in-depth theoretical analysis of our experimental setting is needed. Such analysis is beyond the scope of this study.

At the Heavy Ion Accelerator Facility (HIAF) at the Australian National University (ANU), we explored whether the observed



**Figure 3 | Spectra used to determine the NEEC probability in  $^{93}\text{Mo}$ .** The spectra were generated by applying double gates on the Doppler-shifted 241-keV peak and either the 1,442- or 686-keV transitions, placed so as to avoid overlap with the 1,478- and 685-keV  $^{93}\text{Mo}$  lines from stopped nuclei at particular angles. **a**, High-energy part of the spectrum, with Doppler correction. **b**, Low-energy part of the same spectrum, but with no Doppler correction. The asterisks mark known transitions in  $^{93}\text{Mo}$ ; the contaminant labelled with a 'c' comes from  $^{95}\text{Ru}$  produced in reactions between the  $^{90}\text{Zr}$  beam and the  $^{12}\text{C}$  layer of the target.

coincidences could have originated from a different reaction in our experiment. A control reaction between  $^7\text{Li}$  and  $^{90}\text{Zr}$  that created  $^{93}\text{Mo}$  recoils with energies that were too small to produce the ionization and electron kinetic energies required for NEEC indeed did not yield the coincident  $\gamma$ -rays that would be a signature of NEEC. Additionally, these  $\gamma$ -rays cannot be attributed to reactions between the  $^{90}\text{Zr}$  projectiles and the C or Pb layers in the target (see Methods). We also calculated the probability of  $^{93\text{m}}\text{Mo}$  being excited to the intermediate state through inelastic scattering (Coulomb and nuclear interaction) at recoil energies above the Coulomb barrier in  $^7\text{Li}$  and in  $^{12}\text{C}$  with the coupled-channel-reaction code FRESKO<sup>25</sup>, and through Coulomb excitation at energies below the barrier in  $^{208}\text{Pb}$  with the semi-classical coupled-channel Coulomb-excitation code GOSIA<sup>26</sup> using RACHEL<sup>27</sup>. The resulting probabilities are  $6 \times 10^{-8}$ ,  $2 \times 10^{-6}$  and  $3 \times 10^{-6}$  in the Li, C and Pb target layers, respectively, all of which are too small to account for the experimental value  $P_{\text{exc}} = 0.010(3)$  deduced from our data. Thus, the observed coincidences that provide experimental evidence of NEEC do not appear to originate from contaminant reactions or from other well established excitation mechanisms.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 26 May; accepted 4 December 2017.**

- Hamilton, J. H. (ed.) *Internal Conversion Processes* 1–13 (Academic Press Inc., 1966).
- Blatt, J. M. & Weisskopf, V. F. *Theoretical Nuclear Physics* 614–622 (John Wiley & Sons, 1952).
- Kibédi, T., Burrows, T. W., Trzhaskovskaya, M. B., Davidson, P. M. & Nestor, C. W. Jr. Evaluation of theoretical conversion coefficients using Brlcc. *Nucl. Instrum. Methods A* **589**, 202–229 (2008).
- Goldanskii, V. I. & Namiot, V. A. On the excitation of isomeric nuclear levels by laser radiation through inverse internal electron conversion. *Phys. Lett. B* **62**, 393–394 (1976).
- Morel, P., Daugas, J. M., Gosselin, G., Méot, V. & Gogny, D. Nuclear excitation by electronic processes: NEEC and NEET effects. *AIP Conf. Proc.* **769**, 1085–1088 (2005).
- Pálffy, A., Evers, J. & Keitel, C. H. Isomer triggering via nuclear excitation by electron capture. *Phys. Rev. Lett.* **99**, 172502 (2007).
- Gosselin, G. & Morel, P. Enhanced nuclear level decay in hot dense plasmas. *Phys. Rev. C* **70**, 064603 (2004).

- Belic, D. *et al.* Photoactivation of  $^{180}\text{Tm}$  and its implications for the nucleosynthesis of nature's rarest naturally occurring isotope. *Phys. Rev. Lett.* **83**, 5242–5245 (1999).
- Carroll, J. J. *et al.* Nuclear structure and depletion of nuclear isomers using electron linacs. *AIP Conf. Proc.* **1525**, 586–594 (2013).
- Roig, O. *et al.* Direct evidence for inelastic neutron “acceleration” by  $^{177}\text{Lu}^{\text{m}}$ . *Phys. Rev. C* **83**, 064617 (2011).
- Karamian, S. A. & Carroll, J. J. Cross section for inelastic neutron “acceleration” by  $^{178}\text{Hf}^{\text{m}2}$ . *Phys. Rev. C* **83**, 024604 (2011).
- Stefanescu, I. *et al.* Coulomb excitation of  $^{68,70}\text{Cu}$ : first use of postaccelerated isomeric beams. *Phys. Rev. Lett.* **98**, 122701 (2007).
- Karamian, S. A. & Carroll, J. J. Calculated yield of isomer depletion due to NEEC for  $^{93\text{m}}\text{Mo}$  recoils. *Phys. At. Nucl.* **75**, 1362–1367 (2012).
- Baglin, C. M. Nuclear data sheets for  $A=93$ . *Nucl. Data Sheets* **112**, 1163–1389 (2011).
- Schiwietz, G. & Grande, P. L. Improved charge-state formulas. *Nucl. Instrum. Methods B* **175–177**, 125–131 (2001).
- Polasik, M. *et al.* Resonance conditions for  $^{93\text{m}}\text{Mo}$  isomer depletion via nuclear excitation by electron capture in a beam-based scenario. *Phys. Rev. C* **95**, 034312 (2017).
- Lee, I.-Y. The Gammasphere. *Nucl. Phys. A* **520**, c641–c655 (1990).
- Anderson, J. T. *et al.* A digital data acquisition system for the detectors at Gammasphere. In *IEEE Nuclear Science Symposium and Medical Imaging Conference* 1536–1540 (IEEE, 2012).
- Sethi, J. *et al.* Low-lying states near the  $I^{\pi}=6^{+}$  isomer in  $^{108}\text{Ag}$ . *J. Phys. G* **43**, 015103 (2016).
- Kimball, J. C., Bittel, D. & Cue, N. A comment on “nuclear excitation by target electron capture”. *Phys. Lett. A* **152**, 367–370 (1991).
- Yuan, Z.-S. & Kimball, J. C. First-principles calculation of the cross sections for nuclear excitation by electron capture of channeled nuclei. *Phys. Rev. C* **47**, 323–328 (1993).
- Gunst, J., Litvinov, Y. A., Keitel, C. H. & Pálffy, A. Dominant secondary nuclear photoexcitation with the x-ray free-electron laser. *Phys. Rev. Lett.* **112**, 082501 (2014).
- Gunst, J., Wu, Y., Kumar, N., Keitel, C. H. & Pálffy, A. Direct and secondary nuclear excitation with x-ray free-electron lasers. *Phys. Plasmas* **22**, 112706 (2015).
- Wu, Y., Gunst, J., Keitel, C. H. & Pálffy, A. Tailoring laser-generated plasmas for efficient nuclear excitation by electron capture. Preprint at <https://arxiv.org/abs/1708.04826> (2017).
- Thompson, I. J. Coupled reaction channels calculations in nuclear physics. *Comput. Phys. Rep.* **7**, 167–212 (1988).
- Cline, D. *et al.* GOSIA user manual for simulation and analysis of Coulomb excitation experiments, [http://www.pas.rochester.edu/~cline/Gosia/Gosia\\_Manual\\_20120510.pdf](http://www.pas.rochester.edu/~cline/Gosia/Gosia_Manual_20120510.pdf) (2012).
- Hayes, A. B. & Cline, D. RACHEL graphical interface to GOSIA, <https://github.com/adamhayes/Rachel> (2017).

**Acknowledgements** C.J.C. and J.J.C. thank A. D. Ayangeakaa for input on the potential contributions of Coulomb excitations to the background and M. S. Litz and N. R. Pereira for discussions. We also thank J. Rohrer for assistance in setting up the Gammasphere experiment and the ATLAS operations staff for their efforts. This work was initiated under the US Army Research Laboratory (ARL) Director’s Research Initiative, award number DRI-FY14-SE-022. Further support was provided by ARL Cooperative Agreements W911NF-12-2-0019 and W911NF-16-2-0034, the US Department of Energy (DOE), Office of Science, Office of Nuclear Physics under contract number DE-AC02-06CH11357, the National Science Foundation under grant number PHY-1203100, the Australian Research Council under grant number FT100100991, and the Polish National Science Centre under grants 2011/01/D/ST2/01286 and 2017/25/B/ST2/00901. M.P., J.R. and A.B.H. received support through Ecopulse, Inc. under ARL contract number W911QX09D0016-0004. This research used resources of Argonne National Laboratory’s ATLAS facility, which is a DOE Office of Science User Facility, and of the HIAF at ANU.

**Author Contributions** C.J.C. led the experimental effort. The experiment was conceptualized by S.A.K. and J.J.C., with the final design provided by C.J.C. and J.J.C. with input from D.J.H. and G.J.L. The targets were prepared by J.P.G. All authors, except S.B., A.B.H. and S.A.K., participated in the experiment. C.J.C. analysed the data, with substantial input from J.J.C. Guidance on the atomic conditions for NEEC was provided by M.P. and J.R. The calculations of inelastic-scattering cross-sections with FRESKO were performed by S.B. and those of Coulomb excitation with GOSIA by A.B.H. We wish to call attention to the role of our late colleague S.A.K., who provided the initial impetus for this work but sadly did not see these results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.J.C. ([christopherj.chiara2.ctr@mail.mil](mailto:christopherj.chiara2.ctr@mail.mil)).

**Reviewer Information** Nature thanks O. Kocharovskaya and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Reaction choice and target construction.** The choice of the beam–target combination for this experiment was based on estimates of the production cross-section of  $^{93\text{m}}\text{Mo}$  compared to other reaction products and by the need for recoil energies to exceed the NEEC threshold. The cross-sections were calculated using the Monte Carlo fusion–evaporation code PACE4<sup>28</sup> and qualitatively verified experimentally at the HIAF at ANU (C.J.C. *et al.*, manuscript in preparation). A solid target was preferred over He gas<sup>13</sup> to provide a more compact setup inside Gammasphere (a gas cell tens of centimetres in length would have been needed) and to strip traversing ions to a higher charge state<sup>15</sup>. The  $^{90}\text{Zr} + ^7\text{Li}$  combination suitably met these criteria.

The construction and handling of the Li target needed to address several considerations. (1) Li readily oxidizes when exposed to air. (2) Because the decay sequence that is indicative of NEEC involves an intermediate state with  $t_{1/2} = 3.5$  ns, enough stopping material must be used to prevent reaction products from moving downstream of the target, out of the view of the Ge detectors. (3) Fusion reactions occurring beyond the first few milligrams per square centimetre of a Li target would produce residues with recoil energies too low to satisfy the NEEC resonance condition. This would yield an incorrect value for the NEEC probability with our method (see Methods section ‘Excitation probability and cross-section’), as the 2,475-keV transition to the isomer could be observed for some events with no possibility of NEEC occurring. Thus, stopping all reaction products with a thick Li target (about  $15 \text{ mg cm}^{-2}$  would be required; <http://www.srim.org>) is not desirable. (4) Beam ions that traverse the Li layer without reacting can also potentially induce reactions, such as fusion–evaporation or deep-inelastic reactions, with any material placed behind the Li as a stopper. Such reactions, particularly deep-inelastic processes, would yield a large background from dozens of different nuclei, potentially clouding the NEEC signature being sought. This contamination can be considerably reduced by ensuring that the energy of the beam ions incident upon the stopper is reduced to below the Coulomb barrier of the stopping material (ideally, an element with large atomic number  $Z$ ). (5) For a given ion kinetic energy, a low- $Z$  stopping material is more effective at stripping electrons from the ions down to the inner atomic shells<sup>15</sup>, a necessary condition for NEEC. (6) The half-lives of most states above the  $21/2^+$  isomer are not known, and some may be comparable to or exceed the stopping time of residues in the target or stopper (of the order of picoseconds). In our approach, observation of a  $\gamma$ -ray directly feeding the isomer was required to ensure that the isomeric state was indeed populated. If the residue has slowed substantially before emission of that  $\gamma$ -ray, the recoil energy may already lie below the NEEC resonance, preventing this process from occurring.

To best address these conflicting requirements, the target was prepared as follows. A  $1.55 \text{ mg cm}^{-2}$  deposit of  $^{\text{nat}}\text{Li}$  (abundance of  $^7\text{Li}$  in  $^{\text{nat}}\text{Li}$ , about 92%) was evaporated onto a  $0.50 \text{ mg cm}^{-2}$   $^{\text{nat}}\text{C}$  foil. For the stopping material,  $33 \text{ mg cm}^{-2}$  of  $^{208}\text{Pb}$  was placed behind a  $4.2 \text{ mg cm}^{-2}$  layer of  $^{\text{nat}}\text{C}$ . The target and stopper were mounted together on a frame with a gap of about 3 mm between the two C layers, as illustrated in Extended Data Fig. 1a. To minimize the risk of oxidation, the target was prepared under vacuum in an evaporator at the ATLAS target laboratory and transported to Gammasphere without breaking the vacuum using the technique described in ref. 29. The target was inserted into the beam path at an angle of  $27^\circ$  from the vertical, increasing the effective thickness of each layer by 12% (Extended Data Fig. 1b). With this target configuration,  $^7\text{Li}(^{90}\text{Zr}, p3n)^{93}\text{Mo}$  reactions in the first layer can produce  $^{93}\text{Mo}$  at high spins and with large recoil energies. The kinetic energy decreases in the remaining Li and thin C layers, before the recoils enter the gap. Then, they drift with constant velocity through vacuum for about 100 ps. Any decays with a cumulative timescale smaller than tens of picoseconds will occur predominantly before the  $^{93}\text{Mo}$  ions reach the thicker C layer across the gap; this ensures that the kinetic energies of  $^{93}\text{Mo}$  recoils in the isomeric state are still above the NEEC threshold. (A target arrangement with a thicker layer of Li evaporated directly onto a  $^{208}\text{Pb}$  backing, with no gap, was used initially. With this target, the line shape of the 2,475-keV  $\gamma$ -ray that feeds the isomer was found to have a sizable narrow component from stopped nuclei. Introducing a gap between the Li target and the stopper eliminates this component, as demonstrated in Extended Data Fig. 2.) The recoils then slow down further in the thicker C layer, where NEEC can occur. Finally, the  $^{93}\text{Mo}$  ions come to rest in the  $^{208}\text{Pb}$  backing, permitting observation of  $\gamma$ -rays emitted from longer-lived states, such as the 3.5-ns intermediate state. The combined thickness of the C layers was chosen to reduce the  $^{90}\text{Zr}$  beam energy to below the Coulomb barrier with  $^{208}\text{Pb}$ , so that only a limited number of well known  $\gamma$ -ray transitions produced through Coulomb excitation of  $^{90}\text{Zr}$  or  $^{208}\text{Pb}$  would be expected.

**Spectrum construction and background subtraction.** The broad range of level half-lives in the reaction products, combined with the arrangement of a target and stopper separated by a vacuum gap, resulted in complex  $\gamma$ -ray energy spectra: prompt  $\gamma$ -rays were emitted by moving nuclei (average velocity in the

gap,  $v = 0.109c$ ), with their energies shifted and peaks broadened owing to the Doppler effect, while slower decays were from stopped nuclei, with no energy shifts or Doppler broadening. Although a correction for the Doppler shift can be applied for a given residue velocity and direction of  $\gamma$ -ray emission, the origin of the  $\gamma$ -rays (from moving or stopped nuclei) is not known on an event-by-event basis. Therefore, it was not possible to apply the appropriate Doppler correction to all  $\gamma$ -rays simultaneously. For the transitions of interest here, the 2,475-keV  $\gamma$ -ray, which lies above the  $21/2^+$  isomer, is Doppler shifted, while the 268-, 685- and 1,478-keV  $\gamma$ -rays all lie below the 3.5-ns state and would appear as narrow, unshifted lines.

The anticipated NEEC signature in this experiment would be the observation of a Doppler-shifted 2,475-keV transition in coincidence with the sequence of the 268-, 685- and 1,478-keV  $\gamma$ -rays from stopped nuclei, bypassing the  $21/2^+$  isomer. Double-gated, background-subtracted energy spectra were generated as follows. A gate at a particular energy, denoted  $g$  in the following, means that the energy of a  $\gamma$ -ray in the detected events fell within designated limits; this gate is composed of the peak of interest ( $p$ ) atop a smooth background ( $b$ ),  $g = p + b$ . A separate energy gate was placed on a nearby, flat part of the spectrum to approximate the background component of the gate. Peak–peak (background-subtracted) coincidence spectra were then constructed for double gates on peaks 1 and 2 using the linear combination

$$p_1 p_2 = (g_1 - b_1)(g_2 - b_2) = g_1 g_2 - g_1 b_2 - b_1 g_2 + b_1 b_2 \quad (1)$$

where each term on the right-hand side represents the energy bins of a spectrum of the  $\gamma$ -rays in coincidence with those falling within the indicated gate and/or background regions.

Although the  $\gamma$ -ray spectrum is sparser around 2.5 MeV than at lower energies, the Doppler-shifted and broadened 2,475-keV peak spans an energy range of approximately 500 keV when observed over all of the Gammasphere detector angles, overlapping with various lines from stopped nuclei at specific angles. For each ring of detectors, the gate and background windows were selected so as to limit inclusion of such overlapping contaminant lines. The gate on the narrow 1,478-keV line from stopped nuclei was considerably simpler, as the peak is at the same energy in all rings. However, for this gate it was necessary to reject Gammasphere rings 6 to 8 ( $70^\circ$  to  $81^\circ$ ) because of interference from the Doppler-shifted 1,442-keV transition at those angles. For each transition (and ring), the width of the background window was made the same as that of the corresponding gate. Extended Data Fig. 3 presents the four component spectra of equation (1) for the double gate on the 2,475-keV and 1,478-keV peaks, which was used for Fig. 2a. The energies in these spectra have not been corrected for the Doppler effect in order to show the lines from the stopped nuclei.

Gating in a nearby energy region to approximate the background under a discrete peak is a well established technique<sup>30</sup>. For the 2,475-keV  $\gamma$ -ray, however, the Doppler broadening of the peak at each angle (full-width at half-maximum of tens of kiloelectronvolts, compared to just a few kiloelectronvolts for lines from stopped nuclei) requires that the background gate be placed farther from the peak position than usual. We therefore investigated the effectiveness of our background subtraction method for the 2,475-keV gate. A close inspection of the spectra in Extended Data Fig. 3 reveals that, for the most part, the same peaks appear in each spectrum, with exceptions including the lines at 123, 203, 770 and 963 keV. These four  $\gamma$ -rays are only visible above the background in spectra  $g_1 g_2$  and  $b_1 g_2$ ; this indicates that they are all in true, prompt coincidence with the 1,478-keV transition (Fig. 1), but they are coincident with only background Compton-continuum  $\gamma$ -rays in the energy region near 2,475 keV. As these  $\gamma$ -rays should not be in true coincidence with the double gate on 2,475-keV and 1,478-keV (even if NEEC occurred), they are considered background lines. To ensure that the background is suitably subtracted, the spectra  $g_1 p_2 = g_1 g_2 - g_1 b_2$  and  $b_1 p_2 = b_1 g_2 - b_1 b_2$  were constructed. The peak areas of the four background  $\gamma$ -rays were fitted in both spectra and the ratio  $k = (\text{area in } g_1 p_2) / (\text{area in } b_1 p_2)$  was calculated for each. The average ratio for the four transitions,  $k_{\text{ave}}$ , essentially describes the amount of background attributable to Compton  $\gamma$ -rays in the 2,475-keV gate that needs to be subtracted to eliminate these peaks, with the resulting spectrum defined as

$$p_1 p_2 = g_1 p_2 - k_{\text{ave}} b_1 p_2 = (g_1 g_2 - g_1 b_2) - k_{\text{ave}} (b_1 g_2 - b_1 b_2) \quad (2)$$

We find  $k_{\text{ave}} = 0.99(8)$ , which means that equations (1) and (2) are equivalent here. The  $p_1 p_2$  spectrum in Fig. 2a, which uses  $k_{\text{ave}} = 0.99$ , demonstrates that the background lines are eliminated whereas the 268- and 685-keV peaks remain. By contrast, eliminating these two lines would require  $k = 1.33(7)$  and  $1.21(7)$ , respectively; if these peaks originated from true coincidence with the 1,478-keV gate and the Compton background around 2,475 keV, their  $k$  values would also be close to  $k_{\text{ave}} = 0.99$ . We used a similar procedure for the double gate on the 2,475- and

268-keV peaks (with a different set of  $\gamma$ -rays in true coincidence with the 268-keV gate) and determined an average ratio  $k_{\text{ave}} = 1.03(10)$ ; the central value 1.03 was used for the spectrum in Fig. 2c.

The method of subtracting spectra gated on energies near the peaks of interest eliminates contributions from the smooth Compton background. We estimate from the  $^{93}\text{Mo}$  reaction rate that potential contributions of chance coincidences (prompt  $\gamma$ -rays from two independent reactions arriving within the same coincidence window) are small, but would be accommodated by a larger value of  $k_{\text{ave}}$  to remove the known background lines; the fact that  $k_{\text{ave}} \approx 1$  suggests that chance coincidences are negligible. A final source of background is the long-lived decay of the  $21/2^+$  isomer, which appears at an effectively constant rate in the microsecond-scale coincidence window; however, since the 268-keV transition does not appear in the natural decay of the 6.85-h isomer, these random coincidences do not interfere with the results.

**Excitation probability and cross-section.** For the background-subtracted spectra presented in Fig. 3, gates were placed on the Doppler-shifted 241-keV peak and on either the 1,442-keV or the 686-keV peaks in each ring of Gammasphere. We took particular care to avoid overlaps at certain angles between these gates and the 1,478- and 685-keV peaks from stopped nuclei. The Doppler-corrected 2,475-keV peak in Fig. 3a does not have a simple Gaussian shape. Rather than fitting the peak by constraining it to a particular shape, we simply determined the area above the background by taking the total number of counts within the energy range of the peak and subtracting the average background derived from the surrounding region of the spectrum. We obtained an area of  $A_{2475} = 1.54(3) \times 10^4$  counts. In the spectrum of Fig. 3b, in which no Doppler correction was applied, the 268-keV peak is a narrow line from stopped nuclei that could be fitted (along with the three other peaks of the spectrum) with a Gaussian atop a flat background, yielding an area of  $A_{268} = 5.1(16) \times 10^2$  counts. Although we did not measure the absolute detector efficiencies  $\varepsilon_{\gamma}$ , the relative efficiencies of Gammasphere at 2,475 keV and 268 keV (see Methods section ‘Energy and efficiency calibrations’) are sufficient for determining the ratio  $R = \varepsilon_{\gamma,2475}/\varepsilon_{\gamma,268} = 0.28$ . By factoring in the intensity of the unobserved internal-conversion branch, which is  $\alpha = 0.0355$  times the  $\gamma$ -ray intensity of the 268-keV E2 transition (this branch is negligible for the 2,475-keV transition; see ref. 3 and <http://bricc.anu.edu.au>), we calculated the excitation probability as  $P_{\text{exc}} = RA_{268}(1 + \alpha)/A_{2475} = 1.0(3)\%$  for  $^{93}\text{Mo}$  nuclei traversing the C target.

As a comparison, in ref. 13 a probability of 0.01% is calculated for NEEC in  $^{93}\text{Mo}$ , but it is stated that this value could be much higher because certain parameters were not very well known. Although the experimental conditions proposed in that work for the observation of NEEC are similar to those used in our experiment, the excitation probabilities cannot be directly compared because the predicted value is dependent upon NEEC occurring in a specific length of helium gas<sup>13</sup>. Similarly, this sensitivity of the excitation probability to the experimental conditions prevents a reliable comparison of our measurement with other existing theoretical predictions.

The excitation probability  $P_{\text{exc}}$  can be expressed in terms of a cross-section  $\sigma_{\text{exc}} = N_{\text{exc}}/(N_{\text{proj}}n)$  where  $N_{\text{exc}}$  and  $N_{\text{proj}}$  are the numbers of excitations and projectiles, respectively, and  $n$  is the number of target atoms per unit area. The ratio  $N_{\text{exc}}/N_{\text{proj}}$  here is equivalent to  $P_{\text{exc}}$ , while  $n$  can be written as  $n = \rho N_A/A_t$ , where  $\rho$  is the target surface density,  $A_t$  is the mass per mole of the target and  $N_A$  is the Avogadro constant. The full thickness of C in the target had  $\rho \approx 5 \text{ mg cm}^{-2}$  and the average cross-section across this thickness would be 40 b. This implies a much larger peak cross-section for the resonance, but it is worth noting that rather large values have been suggested for other scenarios<sup>31</sup>.

**Alternative reactions and excitation mechanisms.** The same reaction that was examined in this work was investigated at ANU with the beam and target nuclei interchanged and at a similar centre-of-mass energy (C.J.C. *et al.*, manuscript in preparation). Both reactions are expected to populate the same set of excited states. However, in normal kinematics, with a light beam ( $^7\text{Li}$ ) on a heavy target ( $^{90}\text{Zr}$ ), the kinetic energy of the resulting  $^{93}\text{Mo}$  recoil would be too low for NEEC to occur. The analysis of that dataset demonstrated that under those conditions the 2,475-keV  $\gamma$ -ray was not coincident with either the 268-keV or the 1,478-keV transitions in  $^{93}\text{Mo}$  or in any other nuclide produced in the  $^7\text{Li} + ^{90}\text{Zr}$  reaction. Similarly, reactions between the  $^{90}\text{Zr}$  projectiles and the C in our layered target were ruled out as the origin of the observed coincidences by examining the  $^{12}\text{C} + ^{90}\text{Zr}$  reaction (C.J.C. *et al.*, manuscript in preparation). Thus, the observed  $\gamma$ -ray sequence 2,475 keV–268 keV–685 keV–1,478 keV did not originate from contaminant fusion–evaporation reactions involving either Li or C. (As noted earlier, reactions between  $^{90}\text{Zr}$  and the Pb stopper are expected to yield only a couple of  $\gamma$ -rays with well known energies following Coulomb excitation.) Although these measurements were performed at a different facility from the NEEC experiment at ATLAS, the detector array at ANU is sufficiently sensitive to detect such

coincidences had they existed in normal kinematics, given the excitation probability deduced from the ATLAS experiment.

Coulomb excitation and inelastic scattering of  $^{93}\text{Mo}$  are possible alternative mechanisms by which the nucleus could be excited from the isomer to the intermediate state (as was observed for  $^{68}\text{mCu}$  in ref. 12). We used the codes GOSIA<sup>26</sup> and RACHEL<sup>27</sup> to calculate the expected yield of 268-keV  $\gamma$ -rays from the intermediate state as a fraction of the total number of  $^{93}\text{Mo}$  recoils incident on  $^{208}\text{Pb}$ . GOSIA requires electromagnetic transition strengths as input. Although few transition strengths in  $^{93}\text{Mo}$  have been experimentally determined, several could be estimated from the analogous transitions in  $^{92}\text{Mo}$  owing to the close structural similarity of these two isotopes<sup>32</sup>. For the 4.85(9)-keV,  $21/2^+ \rightarrow (17/2)^+$  E2 excitation<sup>14</sup>, a reduced transition strength of  $B(E2) = 72e^2 \text{ fm}^4$ , where  $e$  is the charge of the electron, can be obtained from the shell-model estimate of ref. 33, which is about two times larger than the strength<sup>34</sup>  $B(E2; 8^+ \rightarrow 6^+)$  of the analogous transition measured in  $^{92}\text{Mo}$ . Weaker (or unobserved) transitions between other pairs of states were assumed to have smaller transition rates than those noted above. With these assumptions, we determine the probability of exciting the isomer to the intermediate state via Coulomb excitation in  $^{208}\text{Pb}$  to be approximately  $3 \times 10^{-6}$ . This probability is dominated by the direct transition between the two states, with minimal impact from multistep excitations involving other states, and repeating the calculation with a drastically simplified level scheme (containing only these two states plus those below the isomer) yields comparable results. Additionally, even by increasing the strength  $B(E2; 21/2^+ \rightarrow (17/2)^+)$  by 100 times, which exceeds the recommended upper limit<sup>35</sup>, the resulting probability of  $2 \times 10^{-4}$  is still much lower than the experimental  $P_{\text{exc}} = 0.010(3)$ .

We also used FRESKO<sup>25</sup> to calculate the expected cross-sections for the excitation of  $^{93}\text{Mo}$  to the intermediate state via above-barrier inelastic scattering in Li and in C. Using the GOSIA findings as a starting point, a simplified level scheme with only direct excitation of the intermediate state was assumed to be sufficient. The reduced transition strength of  $72e^2 \text{ fm}^4$  was again used. Extended Data Fig. 4 shows the cross-sections as a function of  $^{93}\text{Mo}$  recoil energy in both target materials, as calculated by FRESKO. Taking the average cross-section in each medium separately, the corresponding probabilities for the  $21/2^+ \rightarrow (17/2)^+$  excitation are  $6 \times 10^{-8}$  in Li, assuming  $^{93}\text{Mo}$  production at the centre of the target, and  $2 \times 10^{-6}$  over the full thickness of C.

We note that GOSIA and FRESKO are well established reaction codes validated by substantial experimental data. While various NEEC models have existed for several decades, there have been no prior non-null experimental results to test their reliability.

**Energy and efficiency calibrations.** We performed both energy and relative-efficiency calibrations of the Gammasphere Ge detectors using  $^{152}\text{Eu}$  and  $^{56}\text{Co}$  calibration sources. Each source was positioned in the target location, at the centre of Gammasphere. The energies and intensities of the strongest  $\gamma$ -rays that are emitted after the decay of these nuclides, which span from 122 keV to 1,408 keV ( $^{152}\text{Eu}$ ) and 847 keV to 3,451 keV ( $^{56}\text{Co}$ ), are well known. We fitted the peak centroids and areas and compared them with these known energies and intensities, respectively, using the RadWare suite of analysis codes (<http://radware.phy.ornl.gov>). The data from both sources were combined for the energy calibration and were simultaneously fitted with a second-order polynomial. The detection efficiencies of the Ge detectors had a more complex energy dependence<sup>36</sup> and the two sources were treated separately, as their absolute activities were not precisely known. We performed a preliminary fit of the  $^{152}\text{Eu}$  lines and then normalized the  $^{56}\text{Co}$  relative efficiencies to those of  $^{152}\text{Eu}$  in the region where the energies for the two sources overlap to obtain an efficiency curve for the full 0.1–3.4-MeV energy range. This procedure provided relative efficiencies, but these are sufficient because our analyses use only efficiency ratios.

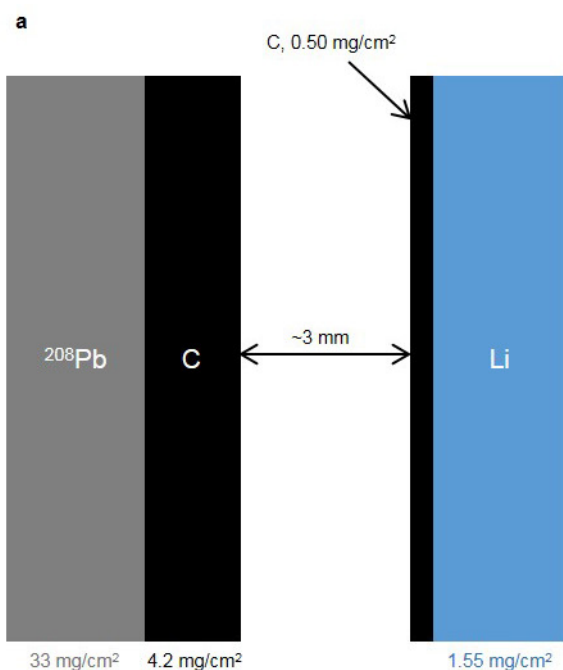
**Uncertainties.** The uncertainty on the counts in each energy bin of a spectrum double-gated on a peak or the background was taken to be the square root of the number of counts in that energy bin. For a background-subtracted spectrum formed by a linear combination of the component spectra (equations (1) or (2)), the uncertainty was that of the individual components added in quadrature. These uncertainties were used as statistical weights for each energy bin in the fits to the peak areas.

The systematic errors associated with the spectra in Fig. 2a, c were taken to be the uncertainties in  $k_{\text{ave}}$ , as determined from the known background  $\gamma$ -rays in each spectrum. We could then define the total uncertainty  $\sigma$  as the statistical and systematic errors added in quadrature. For example, the weighted average of  $k = 1.33(7)$  and  $1.21(7)$  for the 268- and 685-keV peaks in the double gate on the 2,475- and 1,478-keV peaks (Fig. 2a) is  $k_{\text{NEEC}} = 1.27$  with statistical uncertainty 0.05, while the systematic uncertainty on  $k_{\text{ave}} = 0.99$  is 0.08. The total uncertainty is  $\sigma = (0.05^2 + 0.08^2)^{1/2} = 0.09$ , and thus the peaks from NEEC are  $k_{\text{NEEC}} - k_{\text{ave}} = 0.28 = 3\sigma$  above background (compared to a level of about  $7\sigma$  from

a purely statistical uncertainty). A similar analysis for the double gate on the 2,475- and 268-keV peaks (Fig. 2c) also produces a  $3\sigma$  result.

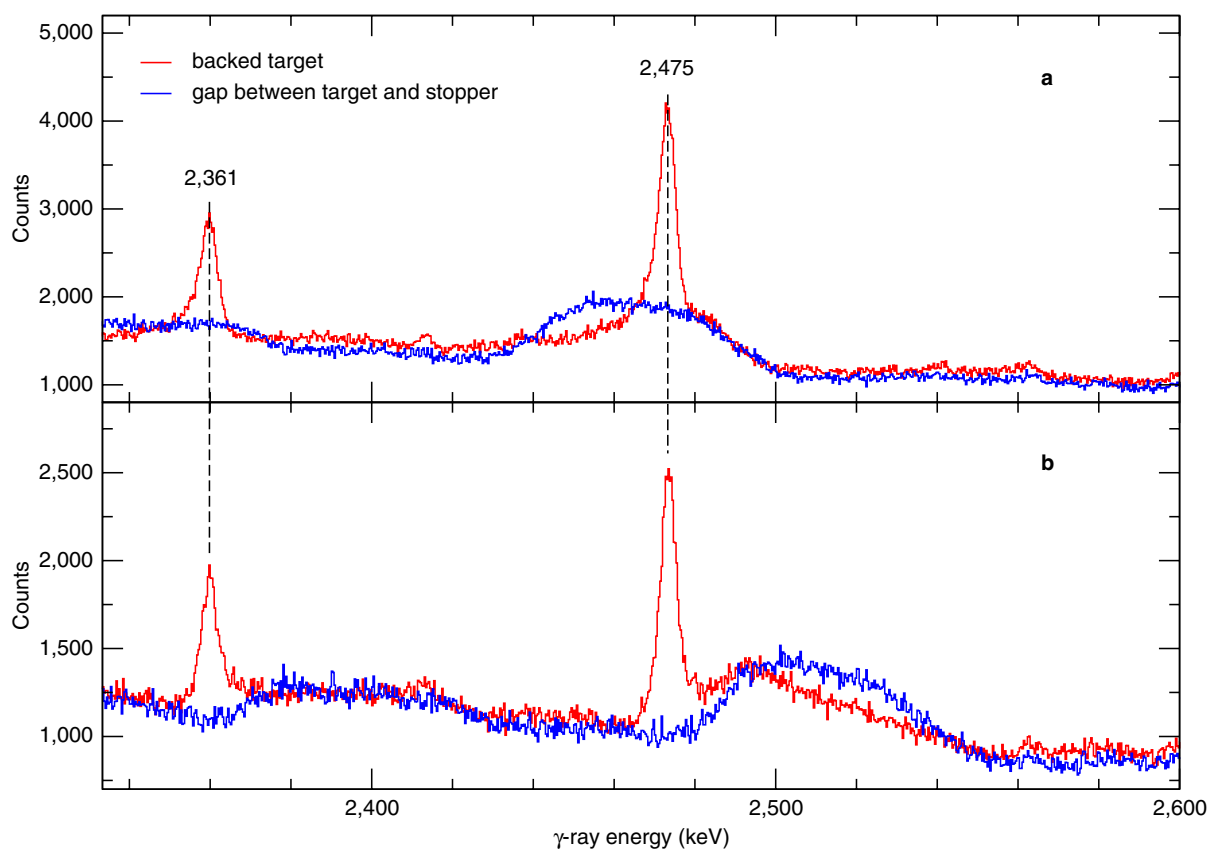
**Code and data availability.** The 7-TB dataset generated and analysed during this study is available from the corresponding author on reasonable request. The codes used in the analysis, aside from those cited in the references, are also available.

28. Tarasov, O. B. & Bazin, D. Development of the program LISE: application to fusion-evaporation. *Nucl. Instrum. Methods B* **204**, 174–178 (2003).
29. McCutchan, E. A., Lister, C. J. & Greene, J. P. A target vacuum interlock system for Gammasphere. *Nucl. Instrum. Methods A* **607**, 564–567 (2009).
30. Radford, D. C. Background subtraction from in-beam HPGe coincidence data sets. *Nucl. Instrum. Methods A* **361**, 306–316 (1995).
31. Pálffy, A., Harman, Z. & Scheid, W. Quantum interference between nuclear excitation by electron capture and radiative recombination. *Phys. Rev. A* **75**, 012709 (2007).
32. Fukuchi, T. *et al.* High-spin isomer in  $^{93}\text{Mo}$ . *Eur. Phys. J. A* **24**, 249–257 (2005).
33. Hasegawa, M., Sun, Y., Tazaki, S., Kaneko, K. & Mizusaki, T. Characteristics of the  $21/2^+$  isomer in  $^{93}\text{Mo}$ : toward the possibility of enhanced nuclear isomer decay. *Phys. Lett. B* **696**, 197–200 (2011).
34. Baglin, C. M. Nuclear data sheets for  $A=92$ . *Nucl. Data Sheets* **113**, 2187–2389 (2012).
35. Firestone, R. B. *et al.* (eds) *Table of Isotopes* 8th edn, Vol. II (John Wiley & Sons, 1996).
36. Radford, D. C. ESCL8R and LEVIT8R: software for interactive graphical analysis of HPGe coincidence data sets. *Nucl. Instrum. Methods A* **361**, 297–305 (1995).



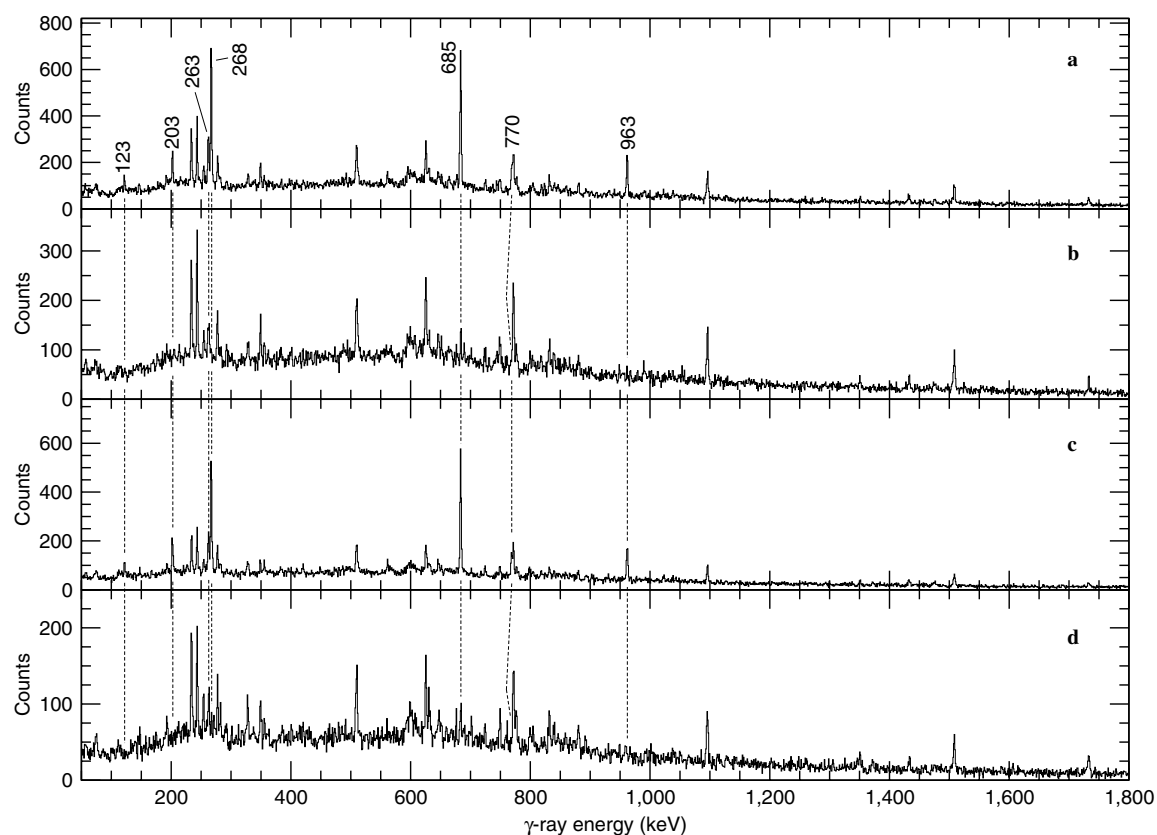
**Extended Data Figure 1 | Configuration of the target used in the experiment.** **a**, Schematic of the target construction showing the layers in which  $^{93}\text{Mo}$  production occurs (Li), NEEC can occur (C) and the backing that stops all recoils ( $^{208}\text{Pb}$ ), in addition to the important gap of about 3 mm that is needed to accommodate the effective half-life for the decay

of the 4,900-keV level. Relative dimensions are not to scale. The beam is incident on the Li surface. **b**, Photograph of the target positioned inside the Gammasphere target chamber. The beam enters from the lower right side and is parallel to the double rods shown in the upper left part of the photograph.



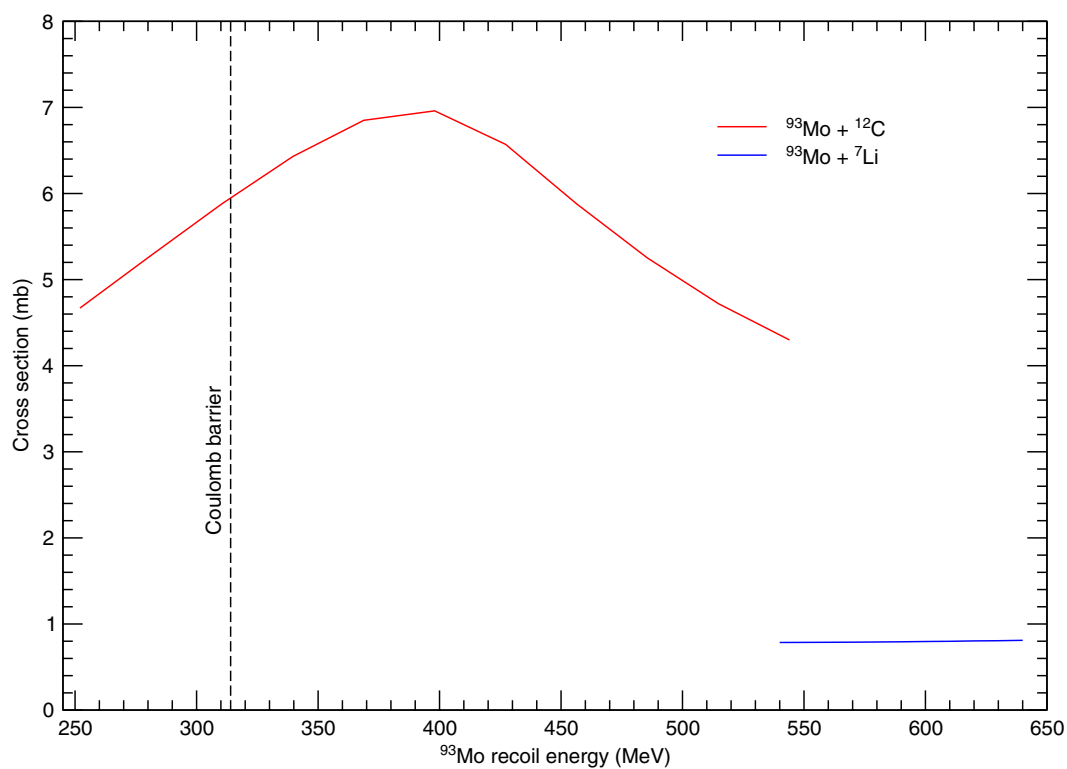
**Extended Data Figure 2 | Spectra showing the line shape of the 2,475-keV transition in  $^{93}\text{Mo}$ .** **a, b,** The spectra are from the detectors in ring 9 of Gammasphere at  $90^\circ$  (**a**) and ring 7 at  $79^\circ$  (**b**). The spectra in red were recorded while using a Li target backed with  $^{208}\text{Pb}$  and with no gap in between. The blue spectra were obtained with a modified target

configuration with a gap of about 3 mm. The 2,361-keV peak corresponds to a transition in  $^{92}\text{Mo}$  that lies below a level with a half-life of  $t_{1/2} = 35$  ps. The similar line shapes of these two transitions support the estimate of a delay of tens of picoseconds in the 2,475-keV emission and therefore the rationale for the final target construction.



**Extended Data Figure 3 | Spectra used to determine background contributions.** Component spectra for the double gate on the Doppler-shifted 2,475-keV  $\gamma$ -ray (1) and the unshifted 1,478-keV  $\gamma$ -ray (2), where gates on the peak and background regions are denoted as 'g' and 'b', respectively (see text). **a**,  $g_1g_2$ . **b**,  $g_1b_2$ . **c**,  $b_1g_2$ . **d**,  $b_1b_2$ . Only those  $\gamma$ -rays in

$^{93}\text{Mo}$  relevant to the discussion are labelled, with the dashed lines marking their energies. We note that the 770-keV peak in **a** is a multiplet with the 773- and 777-keV transitions in  $^{92}\text{Mo}$  and  $^{97}\text{Ru}$ , respectively; only the last two peaks appear in **b** and **d**.



**Extended Data Figure 4 | Calculations of possible competing processes.** The inelastic-scattering cross-sections for exciting  $^{93\text{m}}\text{Mo}$  to the intermediate state, calculated with the code FRESCO, are plotted versus

the energy of recoiling  $^{93}\text{Mo}$  ions traversing the  $^7\text{Li}$  (blue) and  $^{12}\text{C}$  (red) target layers. The initial energy is the average recoil energy corresponding to  $^{93}\text{Mo}$  production at the centre of the Li target.

# Experimental observation of Bethe strings

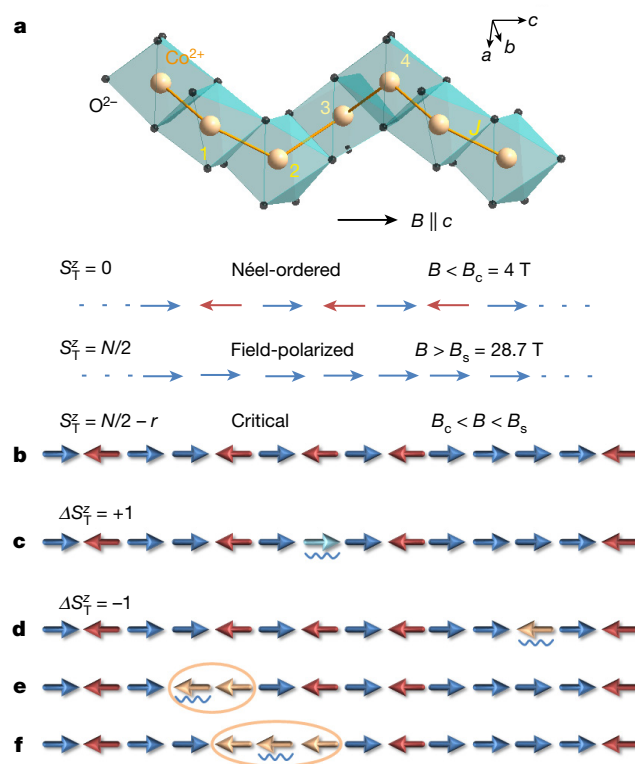
Zhe Wang<sup>1,2</sup>, Jianda Wu<sup>3†</sup>, Wang Yang<sup>3</sup>, Anup Kumar Bera<sup>4,5</sup>, Dmytro Kamenskyi<sup>6</sup>, A. T. M. Nazmul Islam<sup>4</sup>, Shenglong Xu<sup>3</sup>, Joseph Matthew Law<sup>7</sup>, Bella Lake<sup>4,8</sup>, Congjun Wu<sup>3</sup> & Alois Loidl<sup>1</sup>

Almost a century ago, string states—complex bound states of magnetic excitations—were predicted to exist in one-dimensional quantum magnets<sup>1</sup>. However, despite many theoretical studies<sup>2–11</sup>, the experimental realization and identification of string states in a condensed-matter system have yet to be achieved. Here we use high-resolution terahertz spectroscopy to resolve string states in the antiferromagnetic Heisenberg–Ising chain  $\text{SrCo}_2\text{V}_2\text{O}_8$  in strong longitudinal magnetic fields. In the field-induced quantum-critical regime, we identify strings and fractional magnetic excitations that are accurately described by the Bethe ansatz<sup>1,3,4</sup>. Close to quantum criticality, the string excitations govern the quantum spin dynamics, whereas the fractional excitations, which are dominant at low energies, reflect the antiferromagnetic quantum fluctuations. Today, Bethe's result<sup>1</sup> is important not only in the field of quantum magnetism but also more broadly, including in the study of cold atoms and in string theory; hence, we anticipate that our work will shed light on the study of complex many-body systems in general.

Magnons are elementary quasiparticle excitations above the ground state in ferromagnets, which govern the low-temperature thermodynamics<sup>4</sup>. For excited states with two or more magnons, a description in terms of free quasiparticles is very incomplete, especially in one and two dimensions, because the magnons can form bound states that share centre-of-mass momenta owing to the exchange interactions<sup>12</sup>. For a one-dimensional system, bound states of magnons can be viewed as magnetic solitons<sup>13</sup> in the classical limit, which correspond to strings of flipped spins that exist as bound entities in the chain. Studying the dynamical properties of interacting magnetic excitations is of interest not only because of the potential applications in quantum information<sup>14</sup>, but also because it could provide insight into fundamental aspects of quantum magnetism and quantum many-body systems<sup>15</sup>.

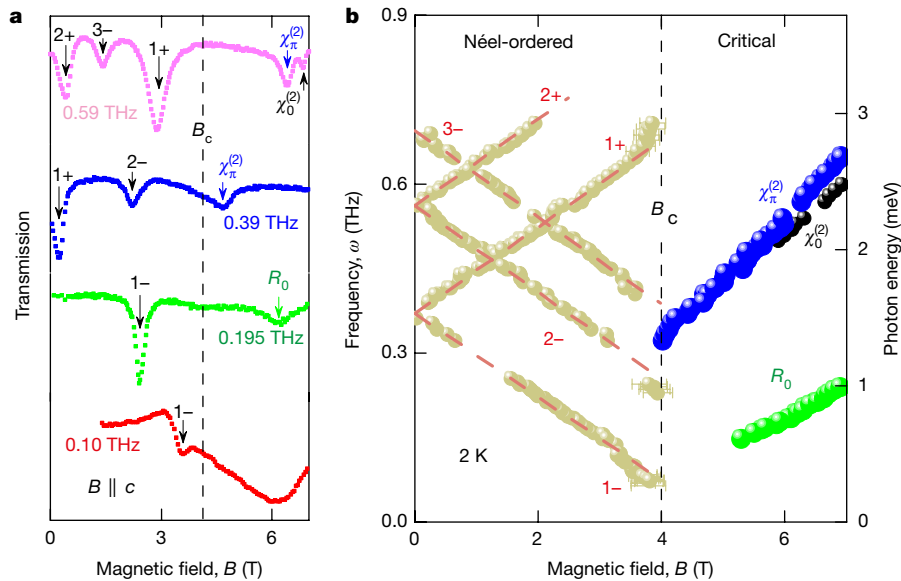
In the one-dimensional spin-1/2 Heisenberg model—a paradigmatic model of interacting spin systems—the existence of bound states was first predicted in the early 1930s by Bethe for two magnons<sup>1</sup>. The systematic ansatz introduced by Bethe for calculating the eigenvalues and eigenstates of the Heisenberg model exactly was later generalized to the description of multi-magnon bound states—the so-called Bethe strings—in models beyond the isotropic limit<sup>3,4,6,8,9</sup>. It is generally believed that spin dynamics is governed by low-energy multi-particle excitations<sup>7,16–19</sup>; however, the excitations of two-magnon bound states (two-string states) have recently been theoretically suggested to be dominant in the isotropic Heisenberg antiferromagnet<sup>10</sup>. Nevertheless, the string excitations are sensitive to exchange anisotropy: for an easy-plane anisotropy, although the spin excitations remain gapless, the dynamical response of string states is substantially smaller compared with the fractional multi-particle excitations (spinons)<sup>8,9</sup>. In addition, in a spin-gapped Heisenberg–Ising antiferromagnet with easy-axis anisotropy, the dynamical properties are dominated by the fractional spinon excitations<sup>19,20</sup>. Hence, experimentally realizing string states is very difficult, and has yet to be achieved in condensed-matter

systems. Here, we perform terahertz spectroscopy on the one-dimensional Heisenberg–Ising antiferromagnetic system  $\text{SrCo}_2\text{V}_2\text{O}_8$  in longitudinal magnetic fields. We show that when the spin gap is closed



**Figure 1 | Quantum spin chain in  $\text{SrCo}_2\text{V}_2\text{O}_8$ , psinon–(anti)psinon pairs and strings.** Psinon–(anti)psinon pairs and strings are the characteristic magnetic excitations in one dimension in the critical regime. **a**, Chain structure of  $\text{SrCo}_2\text{V}_2\text{O}_8$  with a four-fold screw axis along the  $c$  direction. At 1.7 K, Néel-ordered ( $S_T^z = 0$ ) and field-polarized ( $S_T^z = N/2$ ) states are stabilized for longitudinal magnetic fields  $B < B_c = 4$  T and  $B > B_s = 28.7$  T, respectively, where the spins are represented by the arrows. **b**, A representative configuration of the ground state in the critical regime ( $B_c < B < B_s$ ) for total spin- $z$  quantum number  $S_T^z = N/2 - r$  with  $r$  flipped spins with respect to the fully polarized state. **c–f**, Excitations above the ground state are allowed by the selection rules  $\Delta S_T^z = +1$ , yielding psinon–psinon pairs (**c**), and  $\Delta S_T^z = -1$ , yielding psinon–antipsinon pairs (**d**), two-string states (**e**) or three-string states (**f**), which govern the interaction with the magnetic field of a photon. Whereas the psinon–(anti)psinon pairs can propagate throughout the chain without forming bound states, the two-string and three-string states (bound states formed by two and three magnons, respectively; circled) move as entities in the chain. The flipped spin with respect to the ground state (**b**) is indicated by a wiggly line for each excited state (**c–f**).

<sup>1</sup>Experimental Physics V, Center for Electronic Correlations and Magnetism, Institute of Physics, University of Augsburg, 86135 Augsburg, Germany. <sup>2</sup>Institute of Radiation Physics, Helmholtz-Zentrum Dresden-Rossendorf, 01328 Dresden, Germany. <sup>3</sup>Department of Physics, University of California, San Diego, California 92093, USA. <sup>4</sup>Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany. <sup>5</sup>Solid State Physics Division, Bhabha Atomic Research Centre, Mumbai 400085, India. <sup>6</sup>High Field Magnet Laboratory, Radboud University, 6525 ED Nijmegen, The Netherlands. <sup>7</sup>Hochfeld Magnetlab Dresden, Helmholtz-Zentrum Dresden-Rossendorf, 01314 Dresden, Germany. <sup>8</sup>Institut für Festkörperphysik, Technische Universität Berlin, 10623 Berlin, Germany. <sup>†</sup>Present address: Max-Planck-Institut für Physik komplexer Systeme, 01187 Dresden, Germany.



**Figure 2 | Softening of spinons and emergent magnetic excitations at the quantum phase transition in  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** **a**, Transmission spectra of magnetic excitations in  $\text{SrCo}_2\text{V}_2\text{O}_8$  for various frequencies below 1 THz, measured with the applied longitudinal magnetic field  $B \parallel c$  and the electromagnetic wave propagating along the  $c$  axis. Magnetic resonance excitations corresponding to transmission minima are indicated by arrows. **b**, Eigenfrequencies of the resonance modes as a function of the applied magnetic field. For  $B < B_c$ , a series of confined fractional spinon

excitations split in low fields ( $1\pm, 2\pm, \dots$ ) and follow a linear field dependence (dashed lines). Above  $B_c$ , new modes emerge ( $R_0, \chi_0^{(2)}$  and  $\chi_\pi^{(2)}$ ) with completely different field dependencies. Deviations from the linear field dependence appear when approaching the critical field. Error bars indicate the resonance line widths. In both panels, the field-induced phase transition from the Néel-ordered phase to the critical regime is indicated by the vertical dashed line at the critical field  $B_c = 4$  T.

above a field-induced quantum phase transition at  $B_c = 4$  T, many-body two-string and three-string states are identified in the quantum-critical regime before a fully field-polarized state is reached at  $B_s = 28.7$  T. On decreasing the magnetic field from  $B_s$ , the dominant role of the low-energy fractional multi-particles in the dynamical response is gradually taken over by the string states, which govern the quantum spin dynamics close to the quantum phase transition.

The realization of the one-dimensional Heisenberg–Ising model in  $\text{SrCo}_2\text{V}_2\text{O}_8$  is based on its crystal structure and the dominant nearest-neighbour antiferromagnetic interactions (Fig. 1a). The screw chains of  $\text{CoO}_6$  octahedra, with the four-fold screw axis running along the crystallographic  $c$  direction, are arranged in a tetragonal structure. Owing to spin–orbit coupling, the atomic magnetic moments of the  $\text{Co}^{2+}$  ions, comprising the spin and orbital degrees of freedom, are exposed to an Ising anisotropy. The crystal electric field in the  $\text{CoO}_6$  octahedra lifts the twelve-fold degeneracy of the  $\text{Co}^{2+}$  moments, which results in a Kramers doublet ground state with a total angular momentum of  $1/2$  (ref. 21). Magnetization and neutron diffraction experiments reveal that the Ising anisotropy forces the atomic magnetic moments along the  $c$  axis, with Néel-type collinear antiferromagnetic order stabilized below  $T_N = 5$  K (ref. 22).

Superexchange interactions between the magnetic moments in the chains of  $\text{SrCo}_2\text{V}_2\text{O}_8$  are described by the Hamiltonian of the one-dimensional spin-1/2 Heisenberg–Ising model<sup>23,24</sup>:

$$H = J \sum_{n=1}^N [(S_n^x S_{n+1}^x + S_n^y S_{n+1}^y) + \Delta S_n^z S_{n+1}^z] - g_{\parallel} \mu_B B \sum_{n=1}^N S_n^z$$

where  $J > 0$  is the antiferromagnetic coupling between neighbouring spins,  $\Delta > 1$  accounts for the Ising anisotropy between the longitudinal and transverse spin couplings,  $S_n^{x,y,z}$  are the spin components at the  $n$ th site and  $N$  is the length of the chain. The last term corresponds to the Zeeman interaction in a longitudinal magnetic field  $B$  along the  $c$  axis with  $g$ -factor  $g_{\parallel}$  and Bohr magneton  $\mu_B$ .

The Néel ground state at zero field can be illustrated by an antiparallel alignment of neighbouring spins, corresponding to a total spin- $z$  quantum number of  $S_T^z = 0$  (Fig. 1a). One spin flip creates an

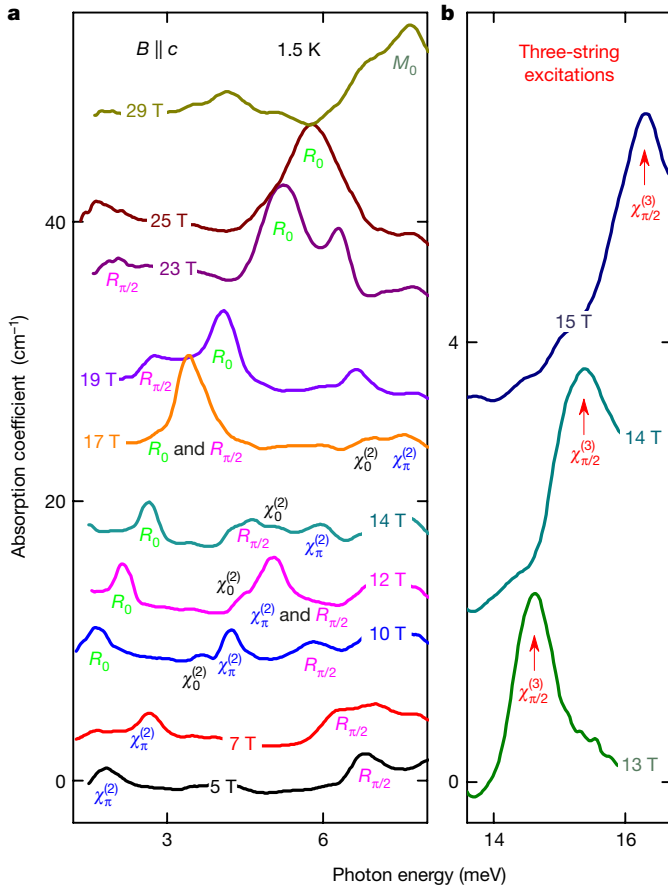
excitation of two spinons<sup>20</sup>, which can propagate separately along the chain via subsequent spin flips. In momentum space the spinons form a two-particle excitation continuum<sup>16–19</sup> that is gapped above the antiferromagnetic ground state.

As a result of the Zeeman interaction, a longitudinal magnetic field can reduce and finally close the spin gap at a critical field  $B_c$  (ref. 2). Before reaching the fully polarized state ( $B > B_s$ ; Fig. 1a) in which the elementary excitations are gapped magnons, the system enters a gapless phase that corresponds to the critical regime ( $B_c < B < B_s$ ). A general ground state in this regime with an arbitrary value of  $S_T^z$  is illustrated in Fig. 1b, and fundamentally new and exotic states can be excited by flipping a single spin (Fig. 1c–f). According to the Bethe ansatz<sup>1</sup>, the spin excitations in the critical regime can be bound states of  $n$  magnons ( $n$ -string states)<sup>3,4</sup> or low-energy spinon-like quasiparticles<sup>5,7–10</sup>. The spinon-like quasiparticles, which form multi-particle continua<sup>25</sup> similarly to spinons (see Methods), are named psinons or antipsinons<sup>7,8</sup> in the context of the Bethe ansatz to distinguish them from the spinons in zero field. We adopt this nomenclature in the following for this reason and, more importantly, because our results reveal that the excitations of psinon–psinon and psinon–antipsinon pairs obey different selection rules (Fig. 1c, d).

Because excitations with  $\Delta S_T^z = +1$  and  $\Delta S_T^z = -1$  are both allowed by the selection rules that govern the interaction with the magnetic field of photons or with the magnetic moment of neutrons, the excited states illustrated in Fig. 1c–f should be observable in optical or neutron-scattering experiments. We use terahertz optical spectroscopy on the Heisenberg–Ising antiferromagnetic chain  $\text{SrCo}_2\text{V}_2\text{O}_8$  in a longitudinal magnetic field up to 30 T. Our results provide clear experimental evidence for the existence of the many-body two-string and three-string states and of the fractional multi-particle excitations that characterize the quantum spin dynamics of the one-dimensional spin-1/2 Heisenberg–Ising model.

In a longitudinal magnetic field,  $S_T^z$  is a good quantum number. The eigenstates of the Heisenberg–Ising model can therefore be classified accordingly, and described by a general Bethe-ansatz wavefunction

$$|\varphi\rangle = \sum_{1 \leq n_1 < \dots < n_r \leq N} a(n_1, n_2, \dots, n_r) |n_1, n_2, \dots, n_r\rangle$$



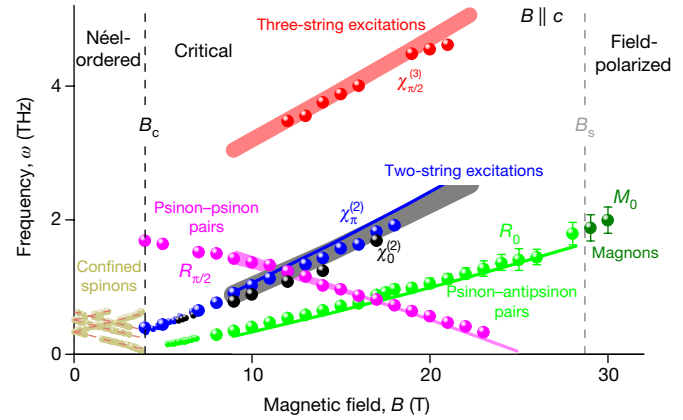
**Figure 3 | Absorption spectra of psinon–psinon, psinon–antipsinon, two-string and three-string excitations for  $B_c < B < B_s$  and of magnons for  $B > B_s$  in  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** **a, b**, Absorption spectra for various longitudinal magnetic fields in the critical regime in the low-energy (**a**) and high-energy (**b**) spectral ranges. **a**, Four types of excitation— $R_0$ ,  $R_{\pi/2}$ ,  $\chi_0^{(2)}$  and  $\chi_{\pi}^{(2)}$ , each with a characteristic field dependence—are observed in the critical regime ( $B_c < B < B_s$ ). Mode  $R_0$  evolves from mode  $M_0$  above  $B_s$ . Whereas mode  $R_{\pi/2}$  softens with increasing fields, the eigenenergies of the other modes increase. **b**, A higher-energy mode ( $\chi_{\pi/2}^{(3)}$ ) can be resolved at relatively low magnetic fields. The spectra are shifted upwards proportional to the corresponding magnetic fields.

for a total spin- $z$  quantum number  $S_T^z = N/2 - r$  (Fig. 1b), with  $N$  denoting the length of the chain and with  $r$  flipped spins (at sites  $n_1, n_2, \dots, n_r$  in the chain) with respect to the fully spin-polarized state  $|\dots \rightarrow \rightarrow \rightarrow \dots\rangle$ , that is,  $|n_1, n_2, \dots, n_r\rangle = S_{n_1}^- S_{n_2}^- \dots S_{n_r}^- |\dots \rightarrow \rightarrow \rightarrow \dots\rangle$ , where  $S_{n_j}^\pm \equiv S_{n_j}^x \pm iS_{n_j}^y$  are the operators that flip the spin of site  $n_j$ . We use the Bethe ansatz to obtain the coefficients  $a(n_1, n_2, \dots, n_r)$  and the eigenenergies of the ground state and the excited states for every  $S_T^z$  (see Methods). The excited states of psinon–(anti)psinon pairs and of  $n$ -strings correspond to the real and complex momenta in the solutions of the Bethe-ansatz equations<sup>1,3,4,6,8–11</sup>, and are labelled as  $R_q$  and  $\chi_q^{(n)}$ , respectively, with the subscript indexing the corresponding transfer momenta. The excitations that are allowed in optical experiments obey the selection rules  $\Delta S_T^z = +1$  or  $\Delta S_T^z = -1$ . These excitations contribute to the dynamic structure factors  $S^{++}(q, \omega)$  or  $S^{+-}(q, \omega)$ , respectively, defined by

$$S^{a\bar{a}}(q, \omega) = \pi \sum_{\mu} \left| \langle \mu | S_q^a | G \rangle \right|^2 \delta(\omega - E_{\mu} + E_G)$$

in which  $\bar{a} = -a$  with  $a \in \{+, -\}$ ,  $|G\rangle$  and  $|\mu\rangle$  are the ground and excited states, with eigenenergies of  $E_G$  and  $E_{\mu}$ , respectively, and

$$S_q^{\pm} = \frac{1}{\sqrt{N}} \sum_n e^{iqn} S_n^{\pm}$$



**Figure 4 | Magnetic excitations in the longitudinal-field Heisenberg–Ising chain  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** Eigenfrequencies are shown as a function of longitudinal magnetic field for all magnetic excitations observed experimentally (symbols). Below  $B_c = 4$  T, confined spinons are observed in the Néel-ordered phase. In the critical regime ( $B_c < B < B_s$ ), excitations of psinon–psinon pairs ( $R_{\pi/2}$  at  $q = \pi/2$ ), psinon–antipsinon pairs ( $R_0$  at  $q = 0$ ), and of complex many-body two-string ( $\chi_0^{(2)}$  and  $\chi_{\pi}^{(2)}$  at  $q = 0$  and  $q = \pi$ , respectively) and three-string ( $\chi_{\pi/2}^{(3)}$  at  $q = \pi/2$ ) bound states are identified by the field dependencies of their eigenfrequencies. Above  $B_s = 28.7$  T, magnons ( $M_0$  at  $q = 0$ ) are observed in the field-polarized ferromagnetic phase. Solid lines display the results of the dynamic structure factors  $S^{++}(q, \omega)$  and  $S^{+-}(q, \omega)$  at  $q = 0$ ,  $q = \pi/2$  and  $q = \pi$  for the corresponding excitations of the one-dimensional spin-1/2 antiferromagnetic Heisenberg–Ising model. For all five modes ( $R_0$ ,  $R_{\pi/2}$ ,  $\chi_0^{(2)}$ ,  $\chi_{\pi}^{(2)}$  and  $\chi_{\pi/2}^{(3)}$ ), excellent agreement between theory and experiment is achieved using an exchange interaction  $J = 3.55$  meV,  $g$ -factor  $g_{\parallel} = 6.2$  and Ising anisotropy  $\Delta = 2$  (ref. 24). Experimental and theoretical line widths are indicated by error bars and shading, respectively.

Hence, we can quantitatively attribute the contributions of string excitations and psinon–(anti)psinon pairs to the relevant dynamic structure factors. The string excitations with higher energies and characteristic field dependencies can readily be distinguished from the low-energy psinon–(anti)psinon pairs (see Methods). This enables us to compare the theoretical calculations to the experimental results precisely as a function of the longitudinal field and to identify the nature of each observed mode.

In Fig. 2a we show transmission spectra at various frequencies below 1 THz as a function of longitudinal magnetic field. At 0.195 THz, two transmission minima are observed, at 2.41 T (mode 1 $-$ ) and 6.18 T (mode  $R_0$ ), below and above the critical field  $B_c = 4$  T. With increasing magnetic fields, mode 1 $-$  shifts to lower frequencies and mode  $R_0$  to higher frequencies. Mode 1 $-$  together with modes 1 $+$ , 2 $-$ , 2 $+$  and 3 $-$ , which are observed at higher frequencies (0.39 THz and 0.59 THz), are known as confined spinon excitations<sup>23</sup> owing to the inter-chain couplings in the gapped Néel-ordered phase ( $B < B_c$ ). Well below the critical field  $B_c$  (Fig. 2b) the confined spinons exhibit Zeeman splitting with linear field dependence<sup>23</sup>. Close to the critical field mode 1 $-$  softens, concomitant with a substantial hardening of mode 1 $+$ . This indicates that the inter-chain couplings are suppressed above  $B_c$  and that the system enters the field-induced critical regime in one dimension. Completely different excitation spectra appear in the critical regime: in the same frequency range (Fig. 2a), we unambiguously observe three sharp modes, denoted by  $R_0$ ,  $\chi_0^{(2)}$  and  $\chi_{\pi}^{(2)}$ . With increasing magnetic field well above the critical field  $B_c$ , the eigenenergies of the three modes increase linearly with different slopes (Fig. 2b).

Using magneto-optic spectroscopy in a high-field laboratory, we are able to extend the search for magnetic excitations to a much larger spectral range and to higher magnetic fields up to 30 T, covering the complete critical regime ( $B_c < B < B_s$ ) and the field-polarized ferromagnetic phase ( $B > B_s = 28.7$  T). As displayed in Fig. 3, the magnetic

excitations are represented by the peaks in the absorption-coefficient spectra at various magnetic fields. At 10 T, we identify not only modes  $R_0$ ,  $\chi_0^{(2)}$  and  $\chi_{\pi/2}^{(2)}$  at 1.66 meV, 3.67 meV and 4.23 meV, respectively, in a sequence of increasing energies, but also a higher-energy mode  $R_{\pi/2}$  at 5.82 meV. Whereas modes  $R_0$  and  $\chi_{\pi/2}^{(2)}$  have comparable absorption coefficients, mode  $\chi_0^{(2)}$  is much weaker, which is consistent with the low-field measurements (such as the 0.59-THz spectrum in Fig. 2a). In higher magnetic fields, mode  $R_{\pi/2}$  softens, whereas the other modes ( $R_0$ ,  $\chi_0^{(2)}$  and  $\chi_{\pi/2}^{(2)}$ ) shift to higher energies. Above the low-energy phonon bands (see Methods), we resolve a further high-energy magnetic excitation  $\chi_{\pi/2}^{(3)}$  of 14.6 meV at 13 T, which shifts to higher energies with increasing magnetic field (Fig. 3b). The field dependence of the eigenfrequencies of the five observed modes ( $R_0$ ,  $R_{\pi/2}$ ,  $\chi_0^{(2)}$ ,  $\chi_{\pi/2}^{(2)}$  and  $\chi_{\pi/2}^{(3)}$ ) is summarized in Fig. 4 (symbols). The field dependencies are linear for all of the modes, each with distinct slope and characteristic energy.

From the Bethe-ansatz calculations<sup>6</sup>, we can single out various excitations and evaluate their respective contributions to the dynamic structure factors  $S^{+-}(q, \omega)$  and  $S^{++}(q, \omega)$  (Methods). In accord with Brillouin-zone folding due to the four-fold screw-axis symmetry of the spin chain<sup>22,24</sup>, in Fig. 4 we show the peak frequencies in  $S^{+-}(q, \omega)$  and  $S^{++}(q, \omega)$  as a function of magnetic field (solid lines) for the transfer momenta  $q = 0$ ,  $q = \pi/2$  and  $q = \pi$ , so as to compare to the terahertz spectroscopic results. Excellent agreement between theory and experiment is achieved for all five distinct magnetic excitations, which enables us to unambiguously identify their nature:  $R_0$  characterizes psinon–antipsinon pairs at  $q = 0$ , whereas  $R_{\pi/2}$  characterizes psinon–psinon pairs at  $q = \pi/2$ . The psinon–antipsinon excitations  $R_0$ , related to the single spin-flip operator  $S_{q=0}^-$ , evolve from the magnon mode  $M_0$  in the field-polarized ferromagnetic phase, in which the largest absorption is observed experimentally (Figs 3a, 4). Most strikingly, we are able to detect and identify the two-string states  $\chi_0^{(2)}$  and  $\chi_{\pi}^{(2)}$  at  $q = 0$  and  $q = \pi$ , respectively, and the three-string states  $\chi_{\pi/2}^{(3)}$  for  $q = \pi/2$ .

The branch of psinon–psinon pairs  $R_{\pi/2}$  belongs to  $S^{++}(q, \omega)$  and obeys the selection rule  $\Delta S_T^z = +1$ . Hence, the psinon–psinon excitations correspond to flipping one spin into the direction of the magnetic field (Fig. 1c), which will decrease the Zeeman energy, so that mode  $R_{\pi/2}$  softens with increasing field. By contrast, the psinon–antipsinon pairs and the string states, which correspond to  $S^{+-}(q, \omega)$ , obey the selection rule  $\Delta S_T^z = -1$  (Fig. 1d–f) and so their eigenenergies increase with magnetic field. The linear dependencies, which arise essentially from the linear dependence of the Zeeman energy on magnetic field, are substantially renormalized as a result of the one-dimensional many-body interactions.

The observation of three-string states reflects a very peculiar feature of the one-dimensional Heisenberg–Ising model: close to quantum criticality, even three magnons can form a stable bound state and, more surprisingly, the bound states of three magnons govern the dynamical response<sup>26</sup>. This is in clear contrast to the isotropic Heisenberg model, in which the two-string states dominate<sup>10</sup>, or to the models with easy-plane anisotropy, in which the fractional multi-particles essentially characterize the dynamical properties<sup>8,9</sup>. Besides their eigenfrequencies, the contribution of the string states to the spin dynamics is also strongly field-dependent<sup>8–10,26</sup>. Starting from the quantum phase transition of the Heisenberg–Ising model<sup>26</sup>, an increase in magnetic field leads to a decreasing contribution of the string excitations. Above the half-saturated magnetization, the low-energy multi-particle excitations become dominant, finally governing the spin dynamics in the fully field-polarized limit (Methods). This is manifested by the rapidly increasing absorption of mode  $R_0$  (see Fig. 3).

We have identified many-body two-string and three-string states in the quantum-critical regime of a one-dimensional spin-1/2 Heisenberg–Ising chain. This represents an example of the experimental realization of strongly correlated quantum states in condensed-matter systems<sup>27</sup>. Further dynamical properties of the string states are expected to be

revealed from inelastic neutron-scattering studies, which can probe the whole Brillouin zone also for the excitation continua<sup>16–19,25</sup>, thus allowing a more detailed comparison to theory. The stability of the string states, as indicated by our results, provides the possibility to study their non-equilibrium behaviour in quantum magnets<sup>28</sup> and cold-atom lattices<sup>29</sup>. Thus, our results pave the way towards the deterministic manipulation of complex magnetic many-body states in solid-state materials and shed light on the study of quantum quench dynamics<sup>11</sup>, the Hubbard model<sup>30</sup>, and string excitations in string theory<sup>15,31</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 July; accepted 28 November 2017.**

- Bethe, H. Zur Theorie der Metalle. I. Eigenwerte und Eigenfunktionen der linearen Atomkette. *Z. Phys.* **71**, 205–226 (1931).
- Yang, C. N. & Yang, C. P. One-dimensional chain of anisotropic spin-spin interactions. II. Properties of the ground-state energy per lattice site for an infinite system. *Phys. Rev.* **150**, 327–339 (1966).
- Gaudin, M. Thermodynamics of the Heisenberg–Ising ring for  $\Delta \geq 1$ . *Phys. Rev. Lett.* **26**, 1301–1304 (1971).
- Takahashi, M. & Suzuki, M. One-dimensional anisotropic Heisenberg model at finite temperatures. *Prog. Theor. Phys.* **48**, 2187–2209 (1972).
- Müller, G., Thomas, H., Beck, H. & Bonner, J. C. Quantum spin dynamics of the antiferromagnetic linear chain in zero and nonzero magnetic field. *Phys. Rev. B* **24**, 1429–1467 (1981).
- Kitanine, N., Maillet, J. M. & Terras, V. Form factors of the XXZ Heisenberg spin-1/2 finite chain. *Nucl. Phys. B* **554**, 647–678 (1999).
- Karbach, M. & Müller, G. Line-shape predictions via Bethe ansatz for the one-dimensional spin-1/2 Heisenberg antiferromagnet in a magnetic field. *Phys. Rev. B* **62**, 14871–14879 (2000).
- Sato, J., Shiroishi, M. & Takahashi, M. Evaluation of dynamic spin structure factor for the spin-1/2 XXZ chain in a magnetic field. *J. Phys. Soc. Jpn* **73**, 3008–3014 (2004).
- Caux, J.-S., Hagemans, R. & Maillet, J. M. Computation of dynamical correlation functions of Heisenberg chains: the gapless anisotropic regime. *J. Stat. Mech.* **2005**, P09003 (2005).
- Kohn, M. Dynamically dominant excitations of string solutions in the spin-1/2 antiferromagnetic Heisenberg chain in a magnetic field. *Phys. Rev. Lett.* **102**, 037203 (2009).
- Ganahl, M., Rabel, E., Essler, F. H. L. & Evertz, H. G. Observation of complex bound states in the spin-1/2 Heisenberg XXZ chain using local quantum quenches. *Phys. Rev. Lett.* **108**, 077206 (2012).
- Wortis, M. Bound states of two spin waves in the Heisenberg ferromagnet. *Phys. Rev.* **132**, 85–97 (1963).
- Fogedby, H. C. The spectrum of the continuous isotropic quantum Heisenberg chain: quantum solitons as magnon bound states. *J. Phys. Chem.* **13**, L195–L200 (1980).
- Subrahmanyam, V. Entanglement dynamics and quantum-state transport in spin chains. *Phys. Rev. A* **69**, 034304 (2004).
- Batchelor, M. T. The Bethe ansatz after 75 years. *Phys. Today* **60**, 36–40 (2007).
- Tennant, D. A., Perring, T. G., Cowley, R. A. & Nagler, S. E. Unbound spinons in the  $S=1/2$  antiferromagnetic chain  $\text{KCuF}_3$ . *Phys. Rev. Lett.* **70**, 4003–4006 (1993).
- Lake, B. *et al.* Confinement of fractional quantum number particles in a condensed-matter system. *Nat. Phys.* **6**, 50–55 (2010).
- Mourigal, M. *et al.* Fractional spinon excitations in the quantum Heisenberg antiferromagnetic chain. *Nat. Phys.* **9**, 435–441 (2013).
- Wu, L. S. *et al.* Orbital-exchange and fractional quantum number excitations in an  $f$ -electron metal,  $\text{Yb}_2\text{Pt}_2\text{Pb}$ . *Science* **352**, 1206–1210 (2016).
- Faddeev, L. D. & Takhtajan, L. A. What is the spin of a spin wave? *Phys. Lett. A* **85**, 375–377 (1981).
- Lines, M. E. Magnetic properties of  $\text{CoCl}_2$  and  $\text{NiCl}_2$ . *Phys. Rev.* **131**, 546–555 (1963).
- Bera, A. K., Lake, B., Stein, W.-D. & Zander, S. Magnetic correlations of the quasi-one-dimensional half-integer spin-chain antiferromagnets  $\text{SrM}_2\text{V}_2\text{O}_8$  ( $M = \text{Co}, \text{Mn}$ ). *Phys. Rev. B* **89**, 094402 (2014).
- Wang, Z. *et al.* Spinon confinement in the one-dimensional Ising-like antiferromagnet  $\text{SrCo}_2\text{V}_2\text{O}_8$ . *Phys. Rev. B* **91**, 140404 (2015).
- Wang, Z. *et al.* From confined spinons to emergent fermions: observation of elementary magnetic excitations in a transverse-field Ising chain. *Phys. Rev. B* **94**, 125130 (2016).
- Stone, M. B. *et al.* Extended quantum critical phase in a magnetized spin-1/2 antiferromagnetic chain. *Phys. Rev. Lett.* **91**, 037205 (2003).
- Yang, W., Wu, J., Xu, S., Wang, Z. & Wu, C. Quantum spin dynamics of the axial antiferromagnetic spin-1/2 XXZ chain in a longitudinal magnetic field. Preprint at [arXiv.org/abs/1702.01854](https://arxiv.org/abs/1702.01854) (2017).
- Essler, F. H. L. & Konik, R. M. In *From Fields to Strings: Circumnavigating Theoretical Physics* (eds Shifman, M. *et al.*) Vol. 1, 684–830 (World Scientific, 2005).

28. Nishida, Y., Kato, Y. & Batista, C. D. Efimov effect in quantum magnets. *Nat. Phys.* **9**, 93–97 (2013).
29. Fukuhara, T. *et al.* Microscopic observation of magnon bound states and their dynamics. *Nature* **502**, 76–79 (2013).
30. Lieb, E. H. & Wu, F. Y. Absence of Mott transition in an exact solution of the short-range, one-band model in one dimension. *Phys. Rev. Lett.* **20**, 1445–1448 (1968).
31. Minahan, J. A. & Zarembo, K. The Bethe-ansatz for  $\mathcal{N}=4$  super Yang-Mills. *J. High Energy Phys.* **3**, 13 (2003).

**Acknowledgements** We thank I. Bloch, M. Karbach, T. Lorenz and X. Zotos for discussions. We acknowledge partial support by the DFG via the Transregional Collaborative Research Center TRR 80, and by the HFML-RU/FOM and the HLD-HZDR, members of the European Magnetic Field Laboratory (EMFL). J.W., W.Y., S.X. and C.W. are supported by NSF grant number DMR-1410375 and AFOSR grant number FA9550-14-1-0168. C.W. also acknowledges partial support from the National Natural Science Foundation of China (grant number 11729402).

**Author Contributions** Z.W. conceived and performed the optical experiments, analysed the data and coordinated the project. J.W., W.Y. and S.X. carried out the Bethe-ansatz calculations. A.K.B. and A.T.M.N.I. prepared and characterized the high-quality single crystals. A.K.B. and J.M.L. performed the high-field magnetization measurements. D.K. assisted with the high-field optical experiments. B.L., C.W. and A.L. supervised the project. Z.W., J.W., W.Y., C.W. and A.L. wrote the manuscript with input from all authors. All authors discussed the results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to Z.W. ([wangzhe.nju@gmail.com](mailto:wangzhe.nju@gmail.com)).

**Reviewer Information** Nature thanks M. Batchelor, J. van den Brink and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Sample preparation.** High-quality single crystals of  $\text{SrCo}_2\text{V}_2\text{O}_8$  were grown using the floating-zone method. Crystal structure and magnetic properties were characterized by X-ray diffraction, neutron diffraction and magnetization measurements<sup>22</sup>. For the optical experiments, single crystals were oriented using X-ray Laue diffraction and cut perpendicular to the tetragonal  $c$  axis with a typical surface area of  $4\text{ mm} \times 4\text{ mm}$  and a thickness of  $1\text{ mm}$ .

**Terahertz spectroscopy in magnetic fields.** Low-frequency optical experiments were carried out in Augsburg. For the transmission spectroscopy below  $1\text{ THz}$ , backward wave oscillators were used as tunable sources of electromagnetic waves. A magneto-optic cryostat (Oxford Instruments/Spectromag) was used to apply external magnetic fields up to  $7\text{ T}$  and to control temperatures down to  $2\text{ K}$ . High-field optical measurements were performed in the High Field Magnet Laboratory in Nijmegen. Transmission spectra were measured using a Fourier-transform spectrometer Bruker IFS-113v, combined with a 30-tesla Bitter electromagnet. Terahertz electromagnetic waves were generated by a Mercury lamp and detected by a silicon bolometer. For all optical measurements, the external magnetic fields were applied parallel to the crystallographic  $c$  axis (longitudinal field  $B \parallel c$ ) and to the propagation direction of the electromagnetic wave (Faraday configuration  $B \parallel k$ ).

**Crystal and magnetic structure of  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** The spin-1/2 Heisenberg–Ising antiferromagnetic chain system  $\text{SrCo}_2\text{V}_2\text{O}_8$  crystallizes in a tetragonal structure with space group  $I4_1cd$  (Extended Data Fig. 1) and lattice constants  $a = 12.2710(1)\text{ \AA}$  and  $c = 8.4192(1)\text{ \AA}$  at room temperature. The screw-chain structure is based on edge-shared  $\text{CoO}_6$  octahedra. The screw axis of every chain is along the crystallographic  $c$  direction with a period of four  $\text{Co}^{2+}$  ions (Extended Data Fig. 1a). In each unit cell, there are four chains, two with left-handed screw axes and two with right-handed ones (Extended Data Fig. 1b). The leading inter-chain couplings are between  $\text{Co}^{2+}$  ions from the neighbouring chains with the same chirality<sup>32,33</sup>, as indicated in Extended Data Fig. 1b. Compared to the intra-chain interaction, the inter-chain coupling is almost negligible,  $J_{\perp}/J < 10^{-2}$  (refs 32, 33). Owing to crystal-field effects and spin–orbit coupling, the atomic magnetic moments of each  $\text{Co}^{2+}$  ion form a ground state of Kramers doublets corresponding to the total angular momentum of  $1/2$ , which comprises spin and orbital degrees of freedom. According to crystal-field theory<sup>21,34</sup> and electron spin-resonance measurements<sup>23</sup>, the magnetic gap between the two lower-lying Kramers doublets is about  $22\text{ meV}$ . The nearest-neighbour exchange interactions between the Kramers doublets in the ground state (corresponding to  $m_J = \pm 1/2$ ) can be described by the spin-1/2 Heisenberg–Ising antiferromagnetic model. Below  $T_N \approx 5\text{ K}$ , a Néel-ordered phase is stabilized in zero magnetic field<sup>22</sup>.

**High-field magnetization in  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** Magnetization measurements were performed at the Dresden High Magnetic Field Laboratory in a pulsed magnetic field up to  $60\text{ T}$ . Extended Data Fig. 2a shows the magnetization in a longitudinal field  $B \parallel c$  at  $1.7\text{ K}$ . The magnetization curve exhibits distinct features in different phases and a clear signature of a quantum phase transition. At zero magnetic field, the Néel-ordered phase is characterized by gapped fractional spinon excitations. A clear onset of magnetization occurs at  $B_c = 4\text{ T}$ , indicating the phase transition to the one-dimensional gapless phase (critical regime). Slightly above  $B_c$ , the magnetization increases quasi-linearly, and then strongly nonlinearly above  $10\text{ T}$ . This behaviour is also reflected in the field derivative of the magnetization,  $dM/dH$ : a nearly constant value is followed by an evident upturn at higher fields (Extended Data Fig. 2b). Field-induced phase transitions at  $B_c$  and  $B_s$  are evidenced by the anomalies in the field derivative of magnetization. Above  $B_s = 28.7\text{ T}$ , the spins are fully field-polarized and the system enters a field-induced ferromagnetic phase with saturated magnetization. We use the spin-1/2 Heisenberg–Ising antiferromagnetic model (equation (1)) to simulate the field dependence of the magnetization in the gapless one-dimensional critical regime between  $B_c$  and  $B_s$  (ref. 2). Using the same parameters as determined from the excitation spectra ( $J = 3.55\text{ meV}$ ,  $g_{\parallel} = 6.2$  and  $\Delta = 2$ ; Fig. 4), we can describe the experimental magnetization curve quite well.

**Low-energy phonon spectrum in  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** The low-energy phonon reflection spectrum was measured using Fourier-transform infrared spectroscopy. Extended Data Fig. 3 shows the phonon spectra of  $\text{SrCo}_2\text{V}_2\text{O}_8$  measured for the polarization  $E^{\omega} \parallel a$  in the relevant spectral range. The strong reflection due to phonon excitations from  $8\text{ meV}$  to  $13.5\text{ meV}$  strongly reduces the transmission at the corresponding spectral range.

**Bethe-ansatz formalism.** We use the Bethe ansatz to obtain the eigenenergies and eigenwavefunctions of the one-dimensional spin-1/2 antiferromagnetic Heisenberg–Ising model:

$$H(\Delta, h) = J \sum_{n=1}^N [(S_n^x S_{n+1}^x + S_n^y S_{n+1}^y) + \Delta S_n^z S_{n+1}^z] - h \sum_{n=1}^N S_n^z \quad (1)$$

in which  $\Delta > 1$  accounts for the Ising anisotropy in  $\text{SrCo}_2\text{V}_2\text{O}_8$  (refs 23, 24) and  $h$  is related to the external magnetic field  $B$  by  $h \equiv g_{\parallel} \mu_B B$ . The Hamiltonian in equation (1) reduces to the isotropic Heisenberg model for  $\Delta = 1$ , whereas for the model with easy-plane anisotropy,  $0 < \Delta < 1$ . The dynamical properties of this model have been studied extensively in various regimes<sup>1–11,26,35–38</sup>. Here, we follow an established theoretical approach<sup>6,35</sup> to solve the Heisenberg–Ising model, and focus on the comparison with the experimental results in  $\text{SrCo}_2\text{V}_2\text{O}_8$ . Details of the theoretical methods and results are presented in ref. 26.

Starting from the fully polarized state as the reference state, we can divide the states into subspaces according to the number of flipped spins  $r$ , or equivalently the total spin in the  $z$  direction  $S_T^z = N/2 - r$ . The corresponding wavefunctions are obtained by solving the Bethe-ansatz equations<sup>35</sup>

$$N\theta_1(\lambda_j) = 2\pi I_j + \sum_{l=1}^r \theta_2(\lambda_j - \lambda_l), \quad j = 1, \dots, r$$

in which  $\theta_1(\lambda)$  and  $\theta_2(\lambda)$ , as functions of rapidity  $\lambda$ , are defined by

$$\theta_n(\lambda) \equiv 2\pi \arctan \left[ \frac{\tan(\lambda)}{\tanh(n\eta/2)} \right] + 2\pi \left[ \frac{\text{Re}(\lambda)}{\pi} + \frac{1}{2} \right]$$

where  $[A]$  denotes the floor function, which gives the greatest integer not larger than  $A$ .

The Bethe quantum numbers  $\{I_j\}$ ,  $j \in \{1, \dots, r\}$ , take integer values when  $r$  is odd and half-integer values when  $r$  is even (see following sections). A state is called a real Bethe eigenstate (psinon or antipsinon) if all of the rapidities  $\lambda_j$  are real and a string state if there is a complex-valued  $\lambda_j$ .

A schematic of the distribution of ground-state Bethe quantum numbers is shown in Extended Data Fig. 4a for  $N = 32$  and  $S_T^z = 8$ . The low-lying excitations of real Bethe states can be classified according to either one of the two patterns that have purely real rapidities  $\{\lambda_j\}$ ,  $j \in \{1, \dots, r\}$  (refs 4, 6–10, 35):  $n$ -pair psinon–psinon states ( $n\psi\psi$ ) and  $n$ -pair psinon–antipsinon states ( $n\psi\psi^*$ ). The Bethe quantum numbers of an  $n\psi\psi$  state can be produced by first extending the left (right) edge of the ground-state Bethe quantum numbers further to the left (right) by  $n$ , then removing  $2n$  numbers within the extended range. Those of an  $n\psi\psi^*$  state can be obtained by removing  $n$  numbers in the ground-state range and putting them outside the range. Illustrations of the two situations are shown in Extended Data Fig. 4b and c for  $1\psi\psi$  and  $1\psi\psi^*$ , respectively. Extended Data Fig. 4d shows the Bethe quantum numbers of a string state, which has larger energy than the real Bethe states.

**Transverse dynamic structure factors.** The dynamic structure factors (DSFs) relevant to the experiment are calculated for Ising anisotropy  $\Delta = 2$  and system size  $N = 200$  using the determinant formulas in the algebraic Bethe-ansatz formalism<sup>6</sup>. Extended Data Fig. 5 displays the results of  $S^{+-}(q, \omega)$  and  $S^{+-}(q, \omega)$  at a representative magnetization  $2m = 0.4$  (ref. 26).  $S^{+-}(q, \omega)$  is dominated by gapless continua that are formed by  $1\psi\psi$  and  $2\psi\psi$  excitations, whereas in  $S^{+-}(q, \omega)$  there are several well-separated dynamical branches. For  $S^{+-}(q, \omega)$ , the lowest-lying gapless continua are formed by  $1\psi\psi^*$  and  $2\psi\psi^*$  excitations. The corresponding spectral weights are located mainly in the energy range  $\hbar\omega < 3J$ . The higher-energy separated continua correspond to two-string and three-string excitations that are found in the spectral ranges  $\hbar\omega > 3J$  and  $\hbar\omega > 5J$ , respectively.

**The momentum-integrated sum rules.** By rescaling  $S^{\pm}$  to  $S^{\pm}/\sqrt{2}$ , the momentum-integrated sum rules can be expressed as

$$R_{a\bar{a}} = \frac{1}{N} \sum_q \int_0^{\infty} \frac{S_{a\bar{a}}(q, \omega)}{2\pi} d\omega = \frac{1}{4} + \frac{m}{2} c_a$$

where  $c_a = \pm 1$  for  $a = \pm$ , respectively.

To evaluate the saturation levels of the sum rules, we define the ratio

$$\nu_{a\bar{a}} = \frac{R'_{a\bar{a}}}{R_{a\bar{a}}}$$

where  $R'_{a\bar{a}}$  is calculated from a partial summation over the selected excitations.

**Ratio of the momentum-integrated intensity of transverse DSFs.** The ratios of momentum-integrated intensity for  $S^{+-}$  and  $S^{+-}$  are displayed in Extended Data Fig. 6a and b, respectively. The calculations are carried out for  $2m$  varied from  $0.1$  to  $0.9$  with steps of  $0.1$  (ref. 26). The good saturations of sum rules (above 87%) for both  $S^{+-}$  and  $S^{+-}$  over the entire range of magnetization indicate that most of the spectral weights are accounted for in the calculations. Particularly for  $S^{+-}$ , the string excitations, and especially the three-string states, become progressively important when magnetization is lowered. Hence, the string states have a predominant role in the dynamical properties in the low-magnetization region.

**Comparison between theory and experiment.** With the four-fold screw axis (Extended Data Fig. 1a), the Brillouin zone is folded by a factor of four in  $\text{SrCo}_2\text{V}_2\text{O}_8$ ; thus, as well as those at  $q=0$ , the excitations at  $q=\pi$  and  $q=\pi/2$  should also be considered for comparison to the terahertz experiments. On the basis of the results shown in Extended Data Fig. 5, in the following we present a detailed analysis of the magnetic-field-dependent properties of the various excitations for these momenta.

Extended Data Figs 7–10 show the DSFs for psinon–psinon, psinon–antipsinon, two-string and three-string excitations, respectively, as functions of energy for magnetizations  $2m=0.1$ – $0.9$  and  $q=0$ ,  $q=\pi/2$  and  $q=\pi$ .

As shown in Extended Data Fig. 7, for the psinon–psinon excitations the DSF spectra exhibit very sharp peaks, and the peak positions shift with magnetic field (or magnetization) only for  $q=\pi/2$ . By contrast, at  $q=0$  and  $q=\pi$  the excitations form continua in the whole spectral range ( $2J$ ), and we cannot define peak positions, or the peak positions are almost independent of magnetic field. Thus, we can clearly distinguish the psinon–psinon excitations at  $q=\pi/2$  from those at  $q=0$  and  $q=\pi$ . Moreover, the softening of the  $q=\pi/2$  modes upon increasing magnetic fields is in clear contrast to the hardening of the other types of excitation (see Fig. 4). This is because psinon–psinon excitations correspond to flipping one spin parallel to the magnetic field. Indeed, the psinon–psinon excitations at  $q=\pi/2$  are observed by our terahertz spectroscopy (see Fig. 4).

The psinon–antipsinon excitations exhibit very sharp peaks at  $q=0$  and  $q=\pi/2$ , but form continua at  $q=\pi$  (see Extended Data Fig. 8). At  $q=0$ , we can clearly see the peaks in the whole field range (Extended Data Fig. 8a), whereas for  $q=\pi/2$  peaks are well resolved only above half-magnetization saturation (Extended Data Fig. 8b). For both of the modes, the peak positions shift to higher energy with increasing magnetic field. The psinon–antipsinon excitations at  $q=0$  are observed experimentally for all magnetic fields above  $B_c=4$  T (see Fig. 4), whereas those at  $q=\pi/2$ , which appear above the field corresponding to half-saturated magnetization,  $B_{\text{hs}}=25$  T and at low energies, are yet to be found. The extended continua at  $q=\pi$  are strongly overlapping for different magnetic fields, and so they cannot be resolved by the present magneto-optical spectroscopy.

Extended Data Fig. 9 shows the DSFs of two-string excitations. At  $q=\pi$  the spectra exhibit sharp peaks with well-defined peak positions, whereas at  $q=0$  the peaks are relatively broad. With increasing magnetic fields, these two modes are

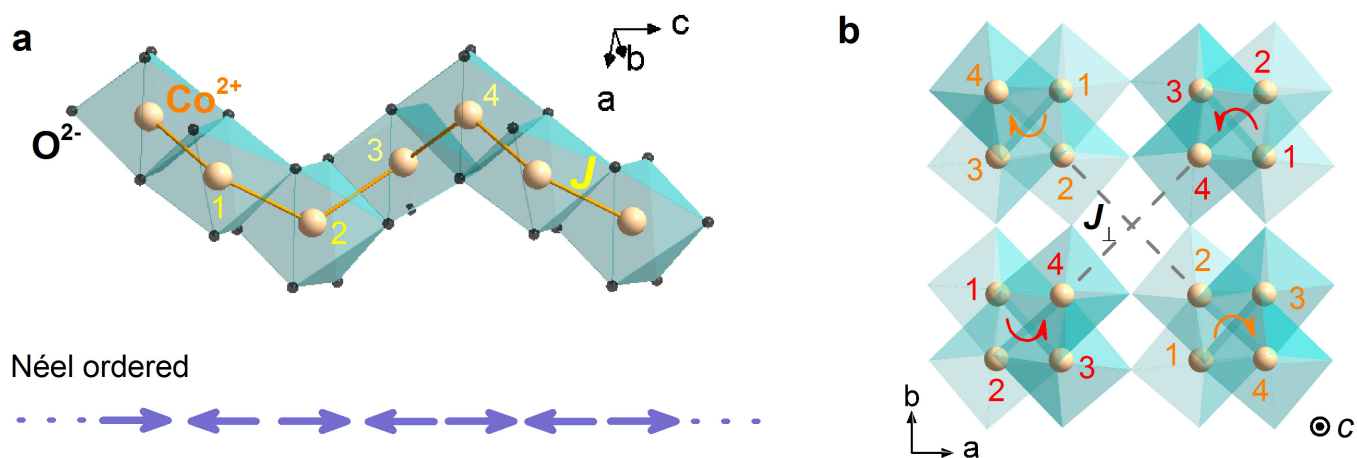
evidently hardening and so can readily be identified. By contrast, the continua at  $q=\pi/2$  overlap with each other for different magnetic fields. Therefore, only at  $q=\pi$  and  $q=0$  can the two-string excitations be observed experimentally (see Fig. 4).

The three-string excitations form extended continua at  $q=0$  and  $q=\pi$ , which overlap for the different magnetic fields, as shown in Extended Data Fig. 10. By contrast, at  $q=\pi/2$  the excitation spectra exhibit well-defined peaks with peak positions shifting to higher energy with increasing magnetic field. This mode can clearly be identified experimentally (see Fig. 4).

To summarize, only the excitations with well-defined peak positions that shift evidently with magnetic fields can be resolved, as have been observed for  $R_0$ ,  $R_{\pi/2}$ ,  $\chi_0^{(2)}$ ,  $\chi_{\pi}^{(2)}$  and  $\chi_{\pi/2}^{(3)}$  by our magneto-optic terahertz spectroscopy (see Fig. 4). By contrast, the other excitations form continua in a broad spectral range and cannot easily be resolved by the terahertz spectroscopy. The theoretical peak positions of the excitations  $R_0$ ,  $R_{\pi/2}$ ,  $\chi_0^{(2)}$ ,  $\chi_{\pi}^{(2)}$  and  $\chi_{\pi/2}^{(3)}$  are plotted as functions of magnetic field and compared to the experimental results. We fix the exchange interaction and Ising anisotropy to the previously determined values for  $\text{SrCo}_2\text{V}_2\text{O}_8$ ,  $J=3.55$  meV and  $\Delta=2$  (ref. 24), and with  $g$ -factor  $g_{\parallel}=6.2$  all five experimentally observed modes can be very well described by the theory, as presented in Fig. 4.

**Data availability.** The data that support the findings of this study are available from the corresponding author on reasonable request.

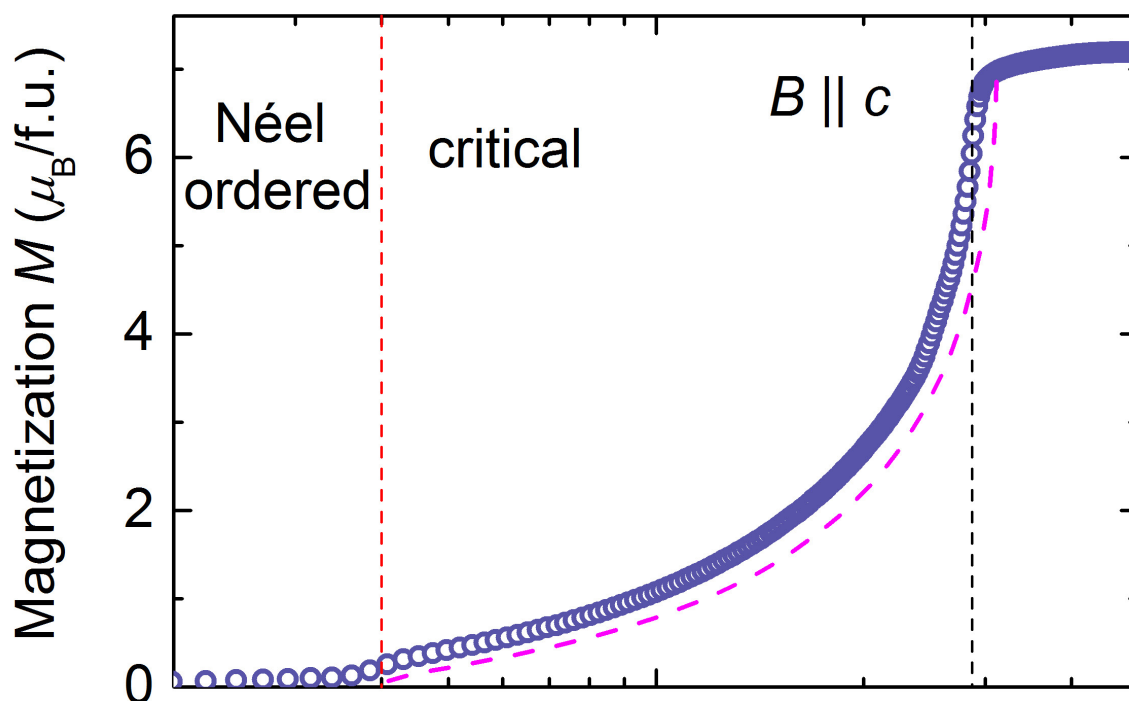
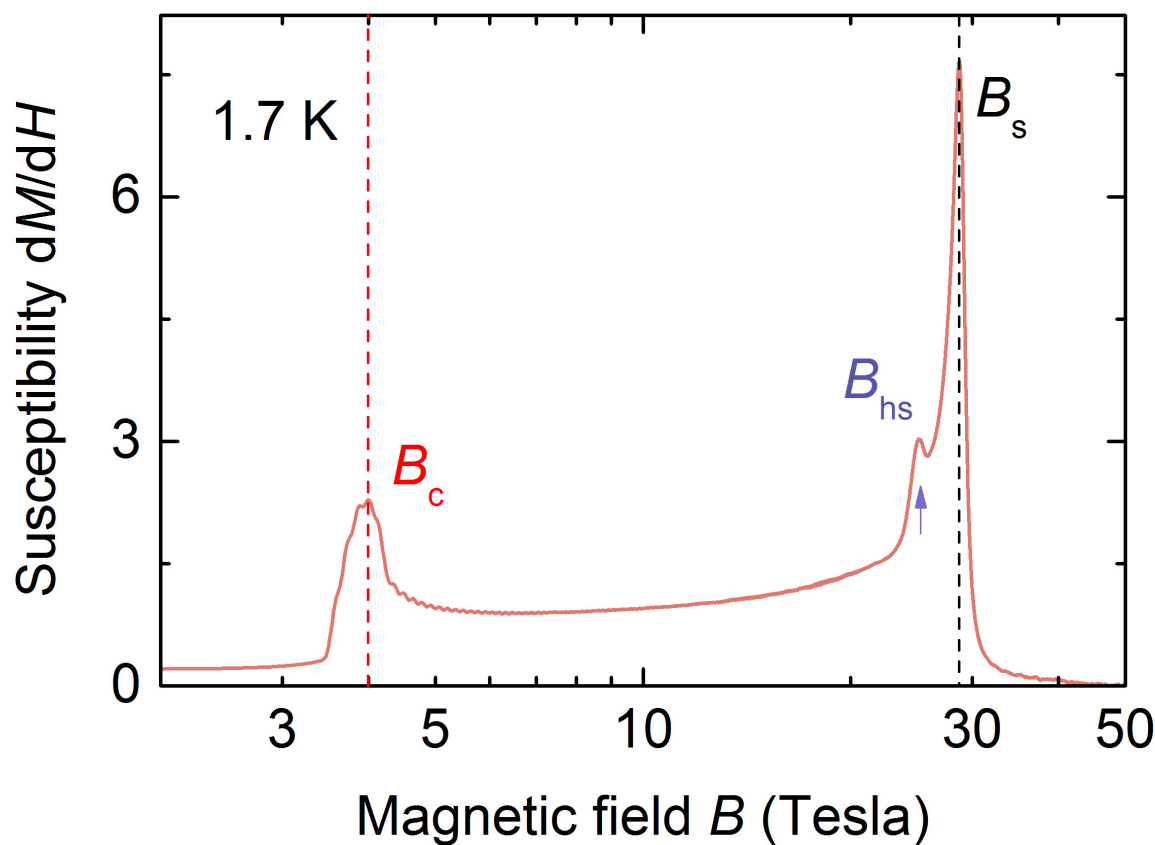
32. Grenier, B. *et al.* Longitudinal and transverse Zeeman ladders in the Ising-like chain antiferromagnet  $\text{BaCo}_2\text{V}_2\text{O}_8$ . *Phys. Rev. Lett.* **114**, 017201 (2015).
33. Bera, A. *et al.* Spinon confinement in a quasi-one dimensional anisotropic Heisenberg magnet. *Phys. Rev. B* **96**, 054423 (2017).
34. Shiba, H., Ueda, Y., Okunishi, K., Kimura, S. & Kindo, K. Exchange interaction via crystal-field excited states and its importance in  $\text{CsCoCl}_3$ . *J. Phys. Soc. Jpn* **72**, 2326–2333 (2003).
35. Takahashi, M. *Thermodynamics of One-Dimensional Solvable Models* (Cambridge Univ. Press, 2005).
36. Caux, J.-S. & Maillet, J. M. Computation of dynamical correlation functions of Heisenberg chains in a magnetic field. *Phys. Rev. Lett.* **95**, 077201 (2005).
37. Pereira, R. G., White, S. R. & Affleck, I. Exact edge singularities and dynamical correlations in spin-1/2 chains. *Phys. Rev. Lett.* **100**, 027206 (2008).
38. He, F., Jiang, Y.-Z., Yu, Y.-C., Lin, H.-Q. & Guan, X.-W. Quantum criticality of spinons. *Phys. Rev. B* **96**, 220401(R) (2017).



**Extended Data Figure 1 | Crystal and magnetic structure of  $\text{SrCo}_2\text{V}_2\text{O}_8$ .**

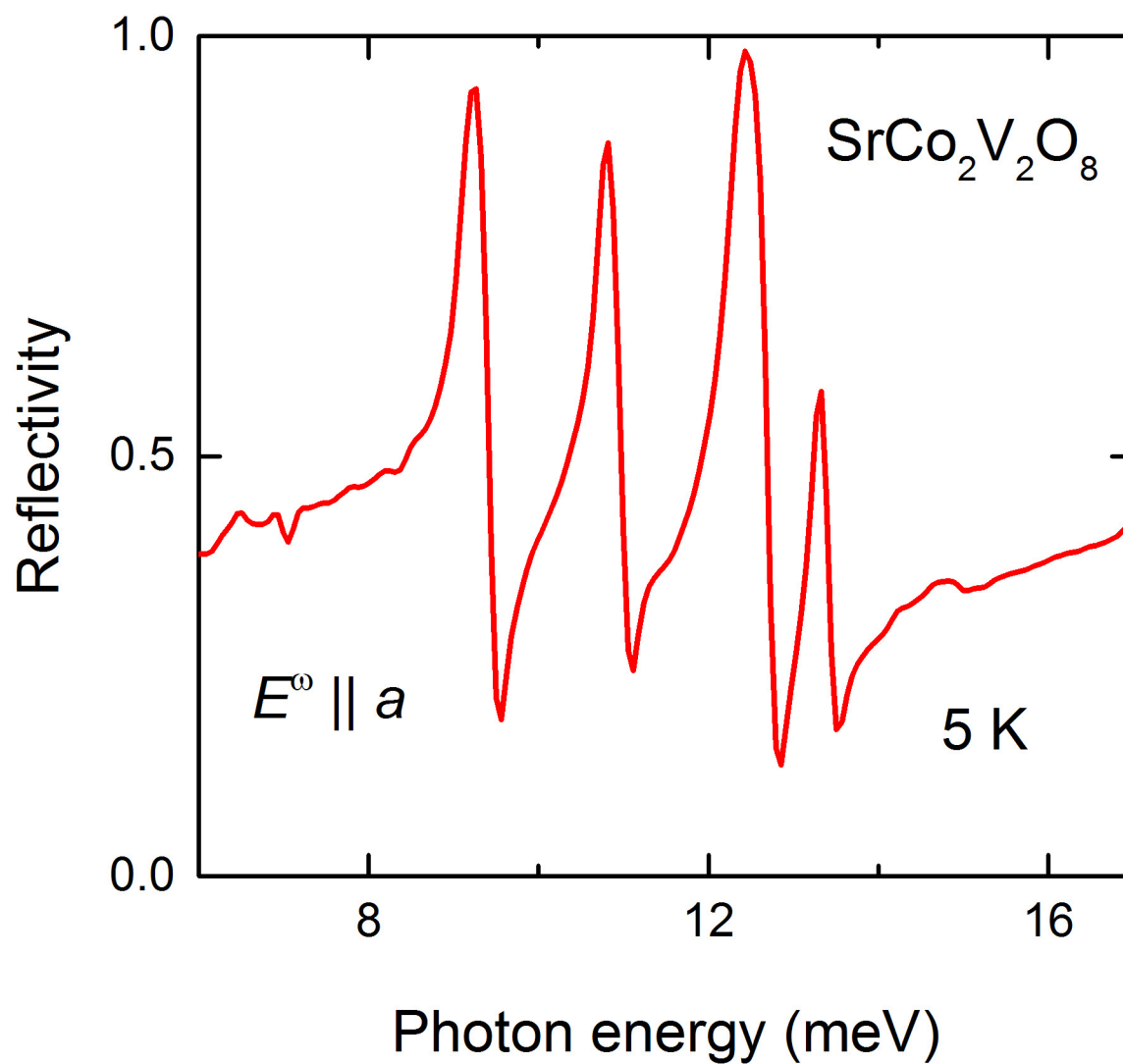
**a**, The screw-chain structure consists of edge-shared  $\text{CoO}_6$  octahedra. Each chain has screw-axis symmetry with a period of four  $\text{Co}^{2+}$  ions (as numbered by the integers 1, 2, 3 and 4), corresponding to the lattice constant along the  $c$  axis. The Néel-ordered phase is illustrated by antiparallel arrows representing magnetic moments at the  $\text{Co}^{2+}$  sites.

Intra-chain nearest-neighbour interaction is denoted by  $J$ . **b**, Viewing from the  $c$  axis, each unit cell contains four screw chains with left- or right-handed screw axes. The leading inter-chain coupling  $J_{\perp}$  is indicated, which is between the  $\text{Co}^{2+}$  ions in the same layer (denoted by the same integer as the Co site) and from chains with the same chirality. It is very small compared to the intra-chain interaction,  $J_{\perp}/J < 10^{-2}$  (refs 32, 33).

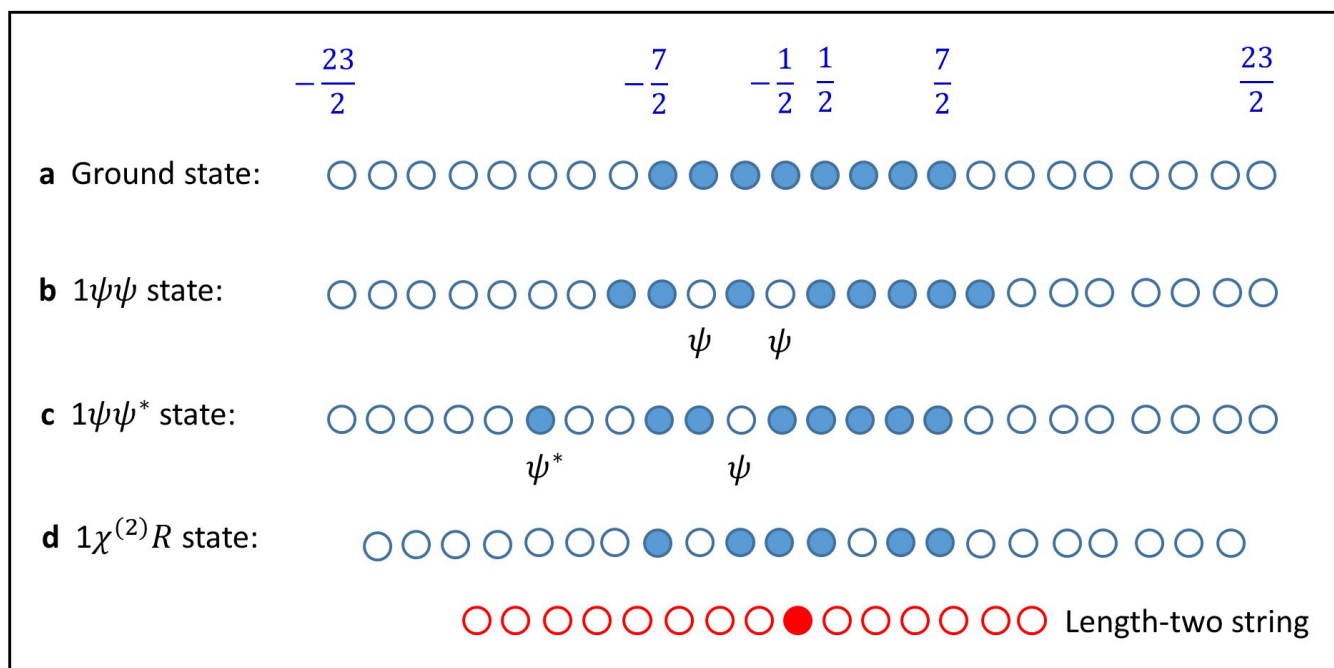
**a****b**

**Extended Data Figure 2 | High-field magnetization and magnetic susceptibility of  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** **a**, Magnetization  $M$  as a function of an applied longitudinal magnetic field  $B$  along the Ising axis ( $B \parallel c$ ), measured at 1.7 K (circles). Theoretical magnetization of the Heisenberg-Ising chain model is shown by the dashed line. **b**, Magnetic susceptibility  $dM/dH$  as a function of the applied longitudinal field  $B$ . A quantum phase

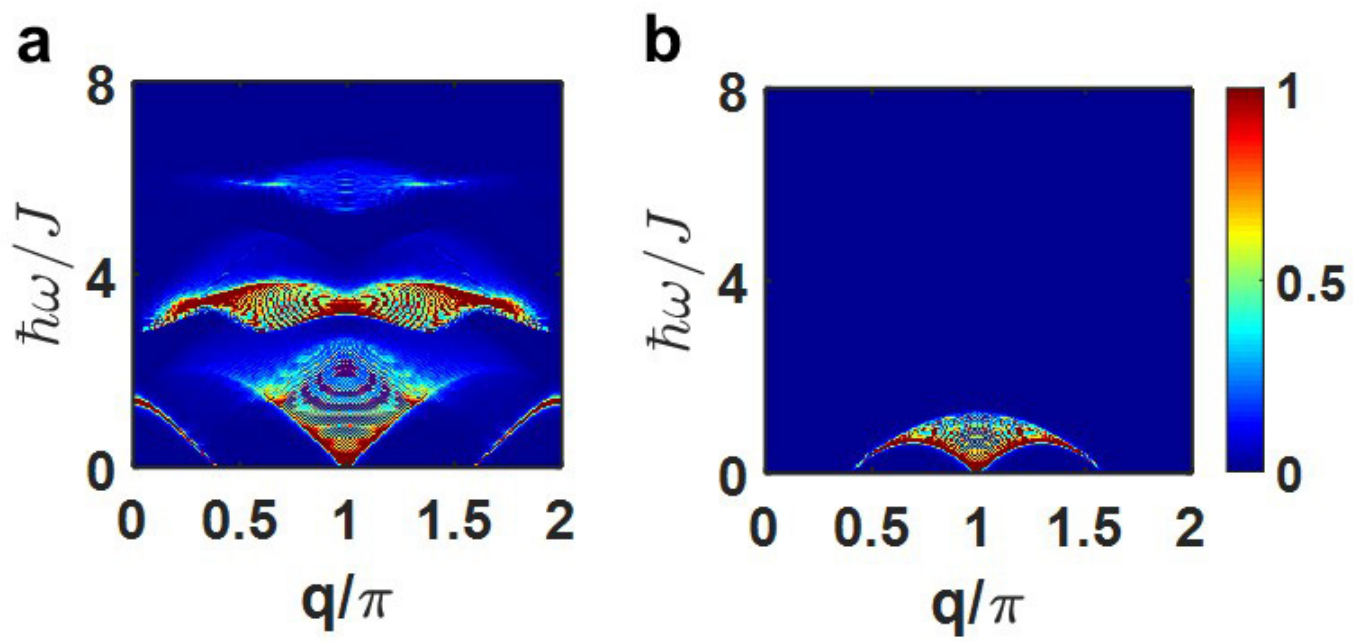
transition from the Néel-ordered phase to the critical phase is revealed by the onset of magnetization and the peak in the susceptibility curve at the critical field  $B_c = 4$  T. Saturated magnetization is observed above the field  $B_s = 28.7$  T and indicated by the sharp peak in the susceptibility. The small anomaly at  $B_{hs} = 25$  T seen in the susceptibility is close to the field of half-saturated magnetization.



**Extended Data Figure 3 | Low-energy phonon spectrum of  $\text{SrCo}_2\text{V}_2\text{O}_8$ .** The phonon spectra of  $\text{SrCo}_2\text{V}_2\text{O}_8$  measured for the polarization  $E^\omega \parallel a$  at 5 K. Strong reflectivity due to phonon excitations is observed in the spectral range 8–13.5 meV.

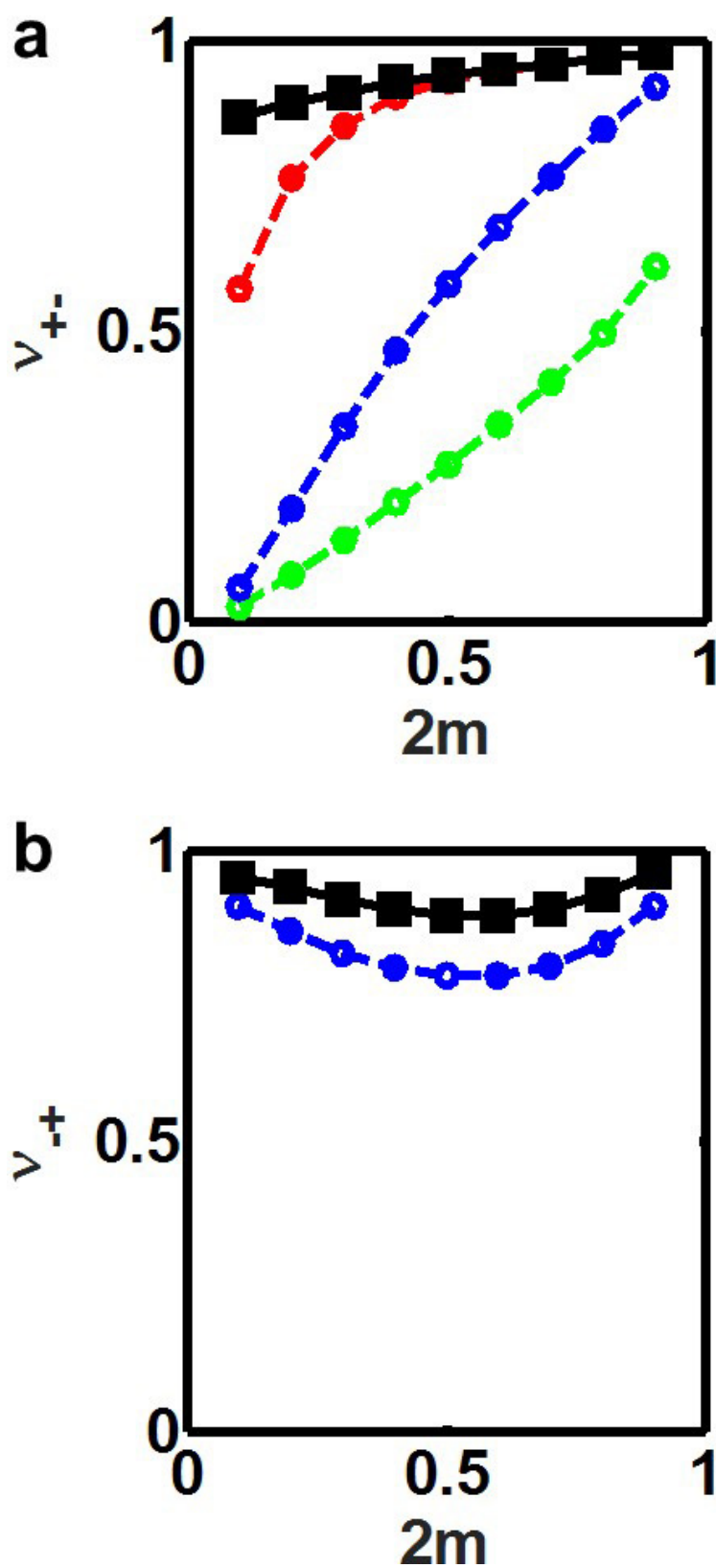


**Extended Data Figure 4 | Schematics of patterns of Bethe quantum numbers.** **a**, The ground state. **b**, One-pair psinon–psinon state  $1\psi\psi$ . **c**, One-pair psinon–antipsinon state  $1\psi\psi^*$ . **d**, Length-two string state  $1\chi^{(2)}R$ . The system size is taken as  $N=32$  and the magnetization is  $S_T^z=8$ .



**Extended Data Figure 5 | DSFs. a, b,  $S^{+-}(q, \omega)$  and  $S^{-+}(q, \omega)$ ,** respectively, as functions of energy  $\hbar\omega/J$  (vertical axis) and momentum  $q/\pi$  (horizontal axis) for  $2m = 0.4$  and  $N = 200$ . The gapless continua are formed by real Bethe eigenstates (psinon–antipsinon pairs in  $S^{+-}$  and

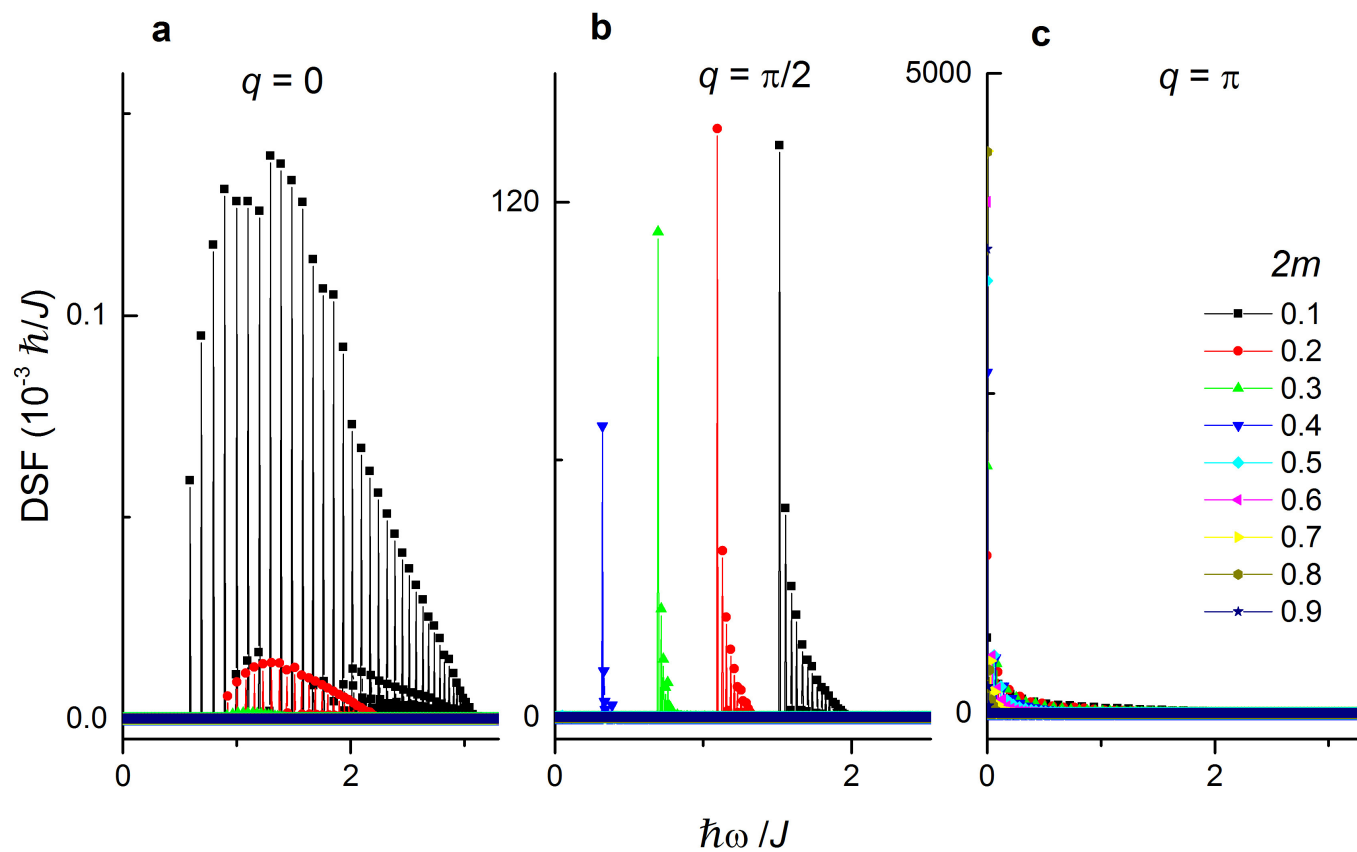
psinon–psinon pairs in  $S^{-+}$ ). For  $S^{+-}$  (a), the higher-energy continua correspond to excitations of two-string ( $\hbar\omega > 3J$ ) and three-string ( $\hbar\omega > 5J$ ) states.



**Extended Data Figure 6 | The momentum-integrated ratios. a, b,  $\nu_{+-}$**  for  $S^{+-}$  and  $\nu_{+-}$  for  $S^{++}$ , respectively, as functions of magnetization  $2m$ . In **a**, the green line is the  $1\psi\psi^*$  contribution. The blue, red and the black

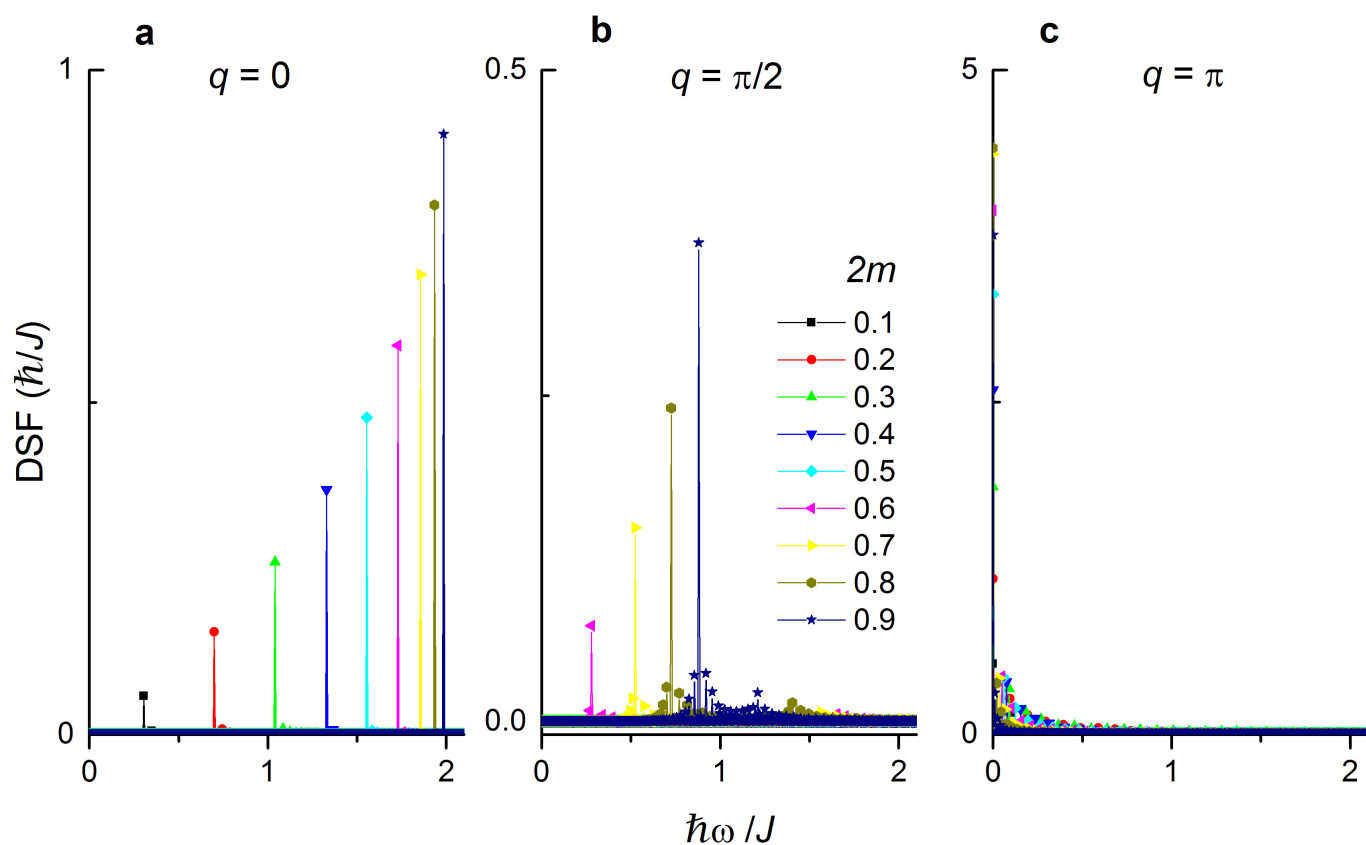
lines are augmented by progressively taking into account the  $2\psi\psi^*$ , two-string and three-string contributions, respectively. In **b**, the blue and black lines represent the  $1\psi\psi$  and  $1\psi\psi + 2\psi\psi$  contributions, respectively.

## psinon-psinon pairs

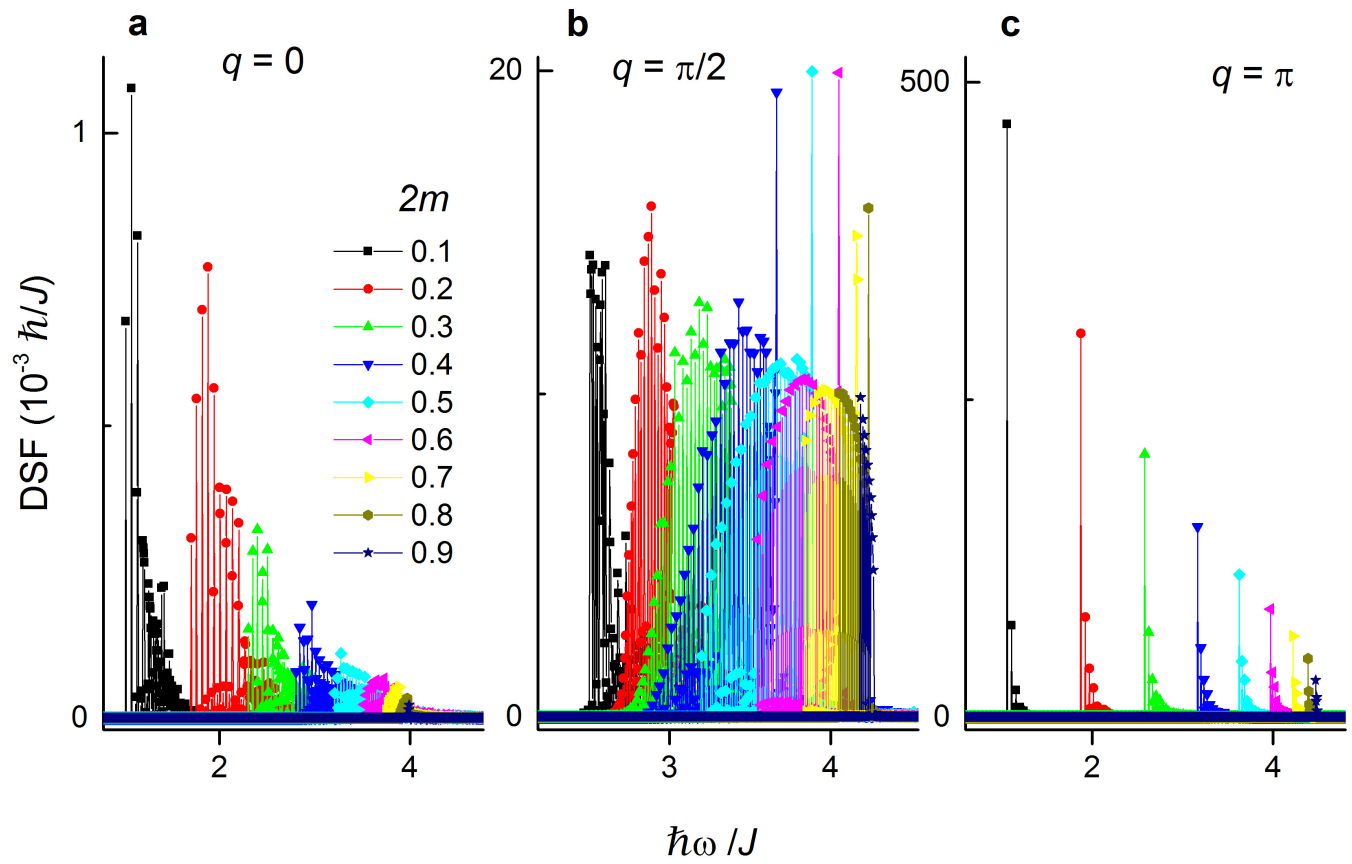


Extended Data Figure 7 | DSF of psinon-psinon pairs as a function of energy for  $2m = 0.1-0.9$ . **a**,  $q = 0$ ; **b**,  $q = \pi/2$ ; **c**,  $q = \pi$ .

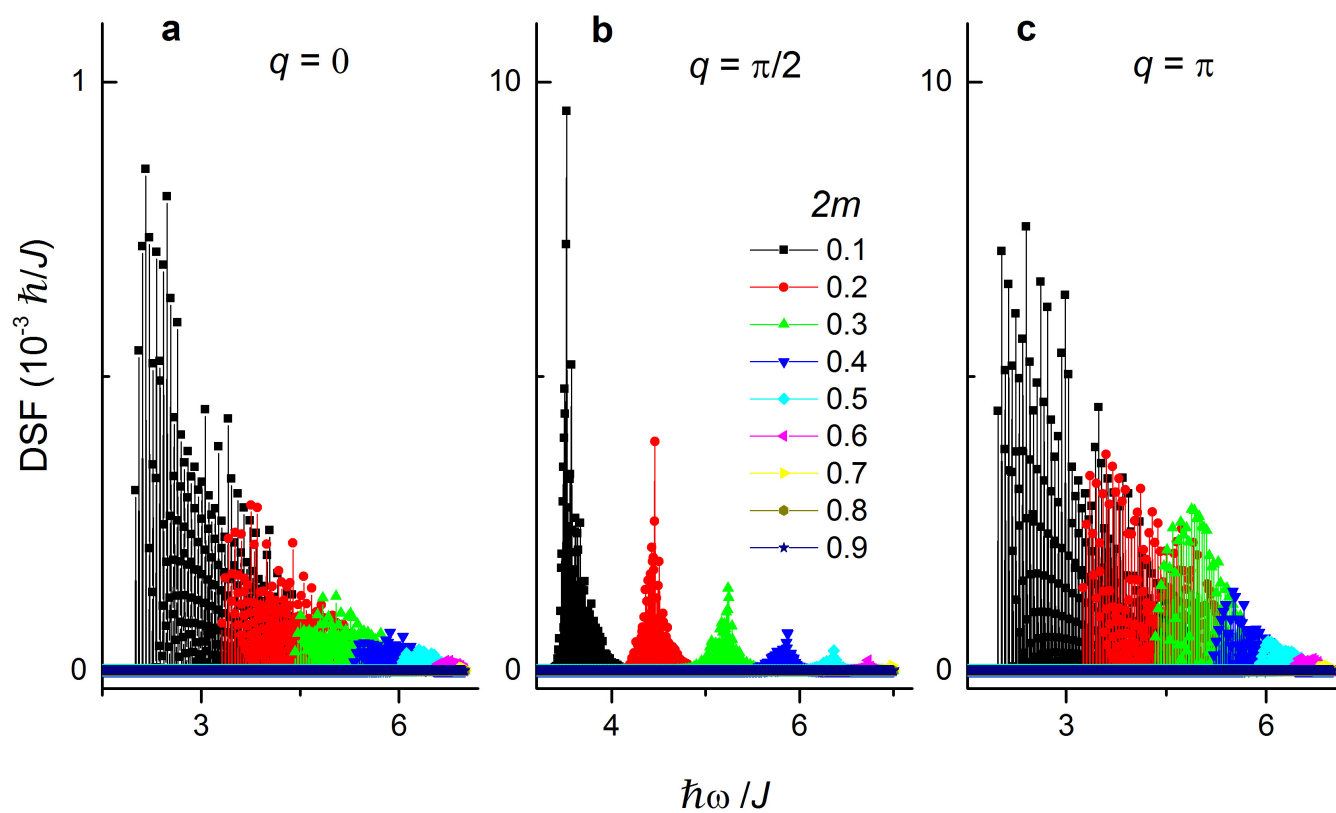
## psinon-antipsinon pairs



Extended Data Figure 8 | DSF of psinon-antipsinon pairs as a function of energy for  $2m = 0.1-0.9$ . **a**,  $q = 0$ ; **b**,  $q = \pi/2$ ; **c**,  $q = \pi$ .

2-string  $\chi_q^{(2)}$ 

Extended Data Figure 9 | DSF factor of two-string states as a function of energy for  $2m = 0.1-0.9$ . **a**,  $q = 0$ ; **b**,  $q = \pi/2$ ; **c**,  $q = \pi$ .

3-string  $\chi_q^{(3)}$ 

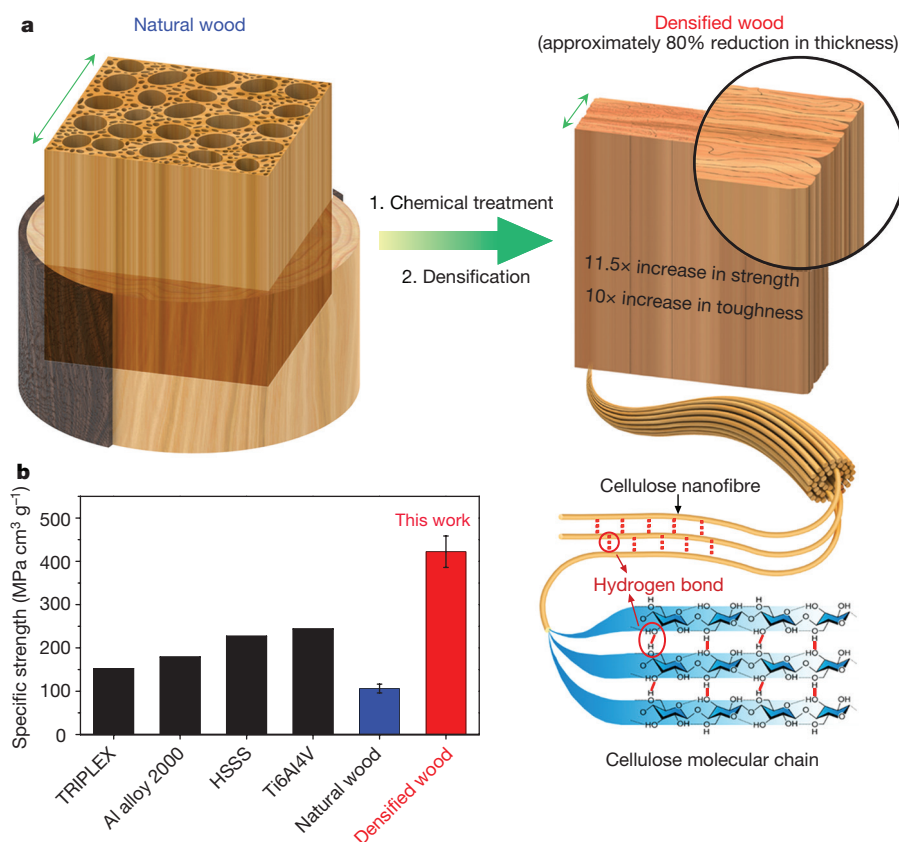
Extended Data Figure 10 | DSF of three-string states as a function of energy for  $2m = 0.1-0.9$ . **a**,  $q = 0$ ; **b**,  $q = \pi/2$ ; **c**,  $q = \pi$ .

# Processing bulk natural wood into a high-performance structural material

Jianwei Song<sup>1\*</sup>, Chaoji Chen<sup>1\*</sup>, Shuze Zhu<sup>2\*</sup>, Mingwei Zhu<sup>1\*</sup>, Jiaqi Dai<sup>1</sup>, Upamanyu Ray<sup>2</sup>, Yiju Li<sup>1</sup>, Yudi Kuang<sup>1</sup>, Yongfeng Li<sup>1</sup>, Nelson Quispe<sup>2</sup>, Yonggang Yao<sup>1</sup>, Amy Gong<sup>1</sup>, Ulrich H. Leiste<sup>3</sup>, Hugh A. Bruck<sup>2</sup>, J. Y. Zhu<sup>4</sup>, Azhar Vellore<sup>5</sup>, Heng Li<sup>6</sup>, Marilyn L. Minus<sup>6</sup>, Zheng Jia<sup>2</sup>, Ashlie Martini<sup>5</sup>, Teng Li<sup>2</sup> & Liangbing Hu<sup>1</sup>

Synthetic structural materials with exceptional mechanical performance suffer from either large weight and adverse environmental impact (for example, steels and alloys) or complex manufacturing processes and thus high cost (for example, polymer-based and biomimetic composites)<sup>1–8</sup>. Natural wood is a low-cost and abundant material and has been used for millennia as a structural material for building and furniture construction<sup>9</sup>. However, the mechanical performance of natural wood (its strength and toughness) is unsatisfactory for many advanced engineering structures and applications. Pre-treatment with steam, heat, ammonia or cold rolling<sup>10–21</sup> followed by densification has led to the enhanced mechanical performance of natural wood. However, the existing methods result in incomplete densification and lack dimensional stability, particularly in response to humid

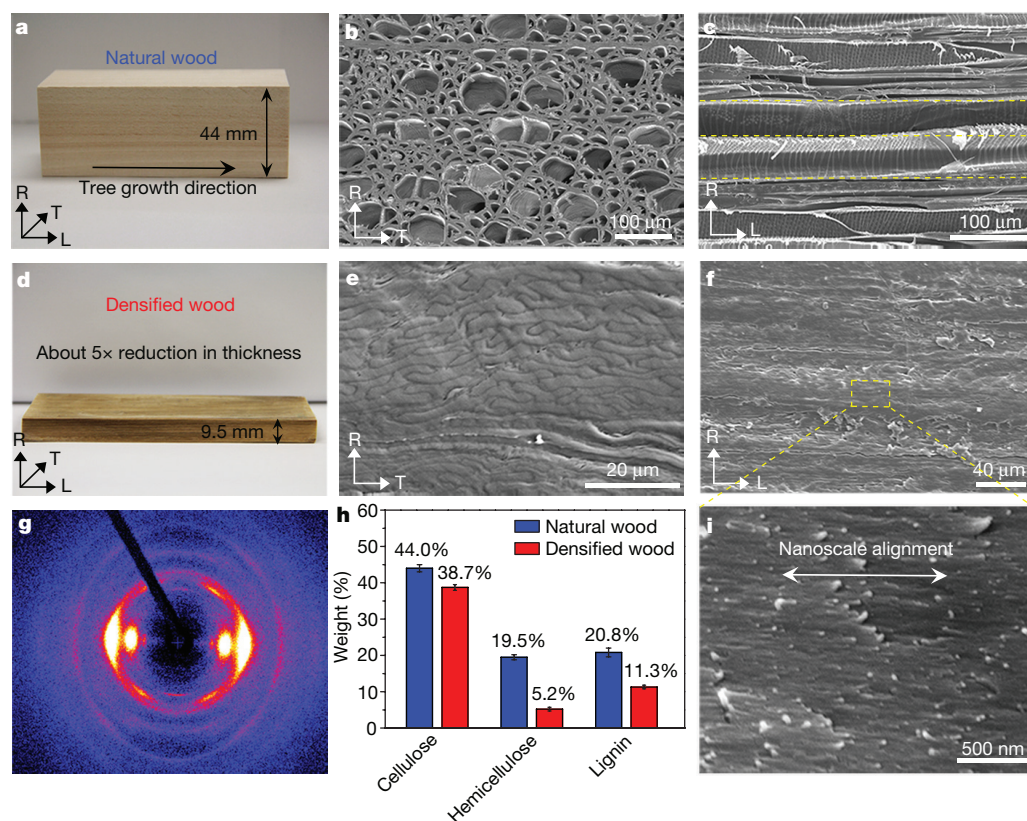
environments<sup>14</sup>, and wood treated in these ways can expand and weaken. Here we report a simple and effective strategy to transform bulk natural wood directly into a high-performance structural material with a more than tenfold increase in strength, toughness and ballistic resistance and with greater dimensional stability. Our two-step process involves the partial removal of lignin and hemicellulose from the natural wood via a boiling process in an aqueous mixture of NaOH and Na<sub>2</sub>SO<sub>3</sub> followed by hot-pressing, leading to the total collapse of cell walls and the complete densification of the natural wood with highly aligned cellulose nanofibres. This strategy is shown to be universally effective for various species of wood. Our processed wood has a specific strength higher than that of most structural metals and alloys, making it a low-cost, high-performance, lightweight alternative.



**Figure 1 | Processing approach and mechanical performance of densified wood.** **a**, Schematic of the top-down two-step approach to transforming bulk natural wood directly into super-strong and tough densified wood. Step 1, chemical treatment to partially remove lignin/hemicellulose; step 2, mechanical hot-pressing at 100 °C, which leads to a reduction in thickness of about 80%. Most of the densified wood consists of well aligned cellulose nanofibres, which greatly enhance hydrogen bond formation among neighbouring nanofibres. **b**, Specific tensile strength of the resulting densified wood ( $422.2 \pm 36.3 \text{ MPa cm}^3 \text{ g}^{-1}$ , mean  $\pm$  standard deviation) is shown to be higher than those of typical metals (the Fe–Al–Mn–C alloy, TRIPLEX and high-specific-strength steel, HSSS), and even of lightweight titanium alloy (Ti6Al4V). Error bars in Figs 1–4 and Extended Data Figs 1–10 show standard deviation with  $n = 5$  repeats, unless noted otherwise.

<sup>1</sup>Department of Materials Science and Engineering, University of Maryland, College Park, Maryland 20742, USA. <sup>2</sup>Department of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA. <sup>3</sup>Department of Aerospace Engineering, University of Maryland, College Park, Maryland 20742, USA. <sup>4</sup>Forest Products Laboratory, USDA Forest Service, Madison, Wisconsin 53726, USA. <sup>5</sup>Department of Mechanical Engineering, University of California Merced, Merced, California 95343, USA. <sup>6</sup>Department of Mechanical and Industrial Engineering, Northeastern University, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.



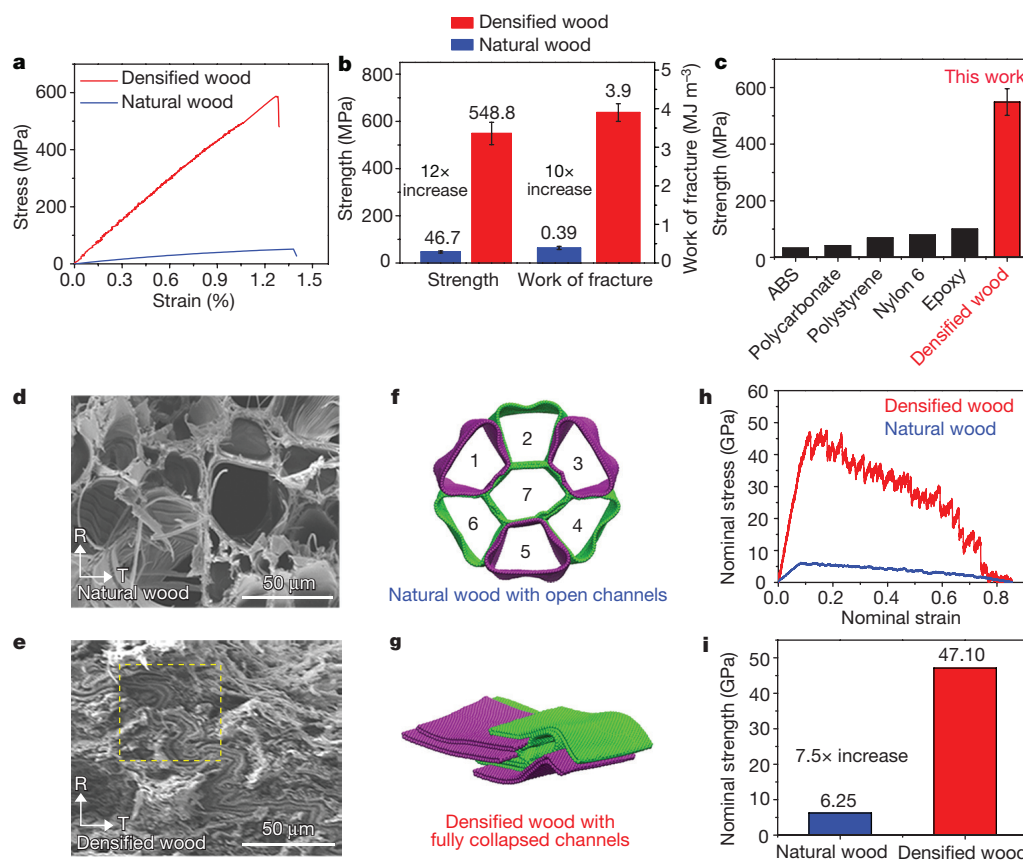
**Figure 2 | Structural characterization of natural wood and densified wood.** **a**, Photograph of natural wood sample. **b**, Scanning electron microscopy (SEM) image of the natural wood sample perpendicular to the tree growth (L) direction, clearly showing the porous structure in the RT plane. **c**, SEM image of the natural wood sample in the RL plane, revealing the cross-section view of the lumina along the L direction (highlighted by dashed lines). **d**, Photograph of densified wood. **e**, SEM image of the densified wood in the RT plane, showing the fully collapsed lumina. The open spaces between the cell walls in natural wood are eliminated, resulting in a unique laminated structure with cell walls tightly

intertwined with each other. **f**, SEM image of the densified wood in the RL plane shows the dense laminated structure cross-section. **g**, Wide-angle X-ray diffraction pattern of the densified wood, showing that the cellulose nanofibre alignment is well preserved after densification. **h**, Chemical treatment leads to substantial removal of lignin (before,  $20.8\% \pm 1.2\%$ ; after,  $11.3\% \pm 0.5\%$ ) and hemicellulose (before,  $19.5\% \pm 0.7\%$ ; after,  $5.2\% \pm 0.5\%$ ), but only modest dissolution of cellulose content (before,  $44.0\% \pm 1.0\%$ ; after,  $38.7\% \pm 0.8\%$ ). **i**, Magnified SEM image of the densified wood, showing the highly aligned cellulose nanofibres.

Figure 1a shows a schematic of our top-down two-step approach to directly transforming bulk natural wood. Our approach involves partial removal of lignin/hemicellulose from bulk natural wood followed by hot-pressing (Fig. 1a; see Methods). Natural wood contains many lumina (tubular channels 20–80  $\mu\text{m}$  in diameter) along the wood growth direction (Fig. 2a–c and Extended Data Fig. 1d, e). Chemical treatment leads to substantial reduction of lignin/hemicellulose content in natural wood, but only modest reduction of cellulose content, largely owing to the different stabilities of these three components in the NaOH/Na<sub>2</sub>SO<sub>3</sub> solution (Fig. 2h). By partial removal of lignin/hemicellulose from the wood cell walls, the wood becomes more porous and less rigid (Extended Data Fig. 1a, b). Upon hot-pressing at 100 °C perpendicular to the wood growth direction, the wood lumina as well as the porous wood cell walls collapse entirely, resulting in a densified piece of wood reduced in thickness to about 20% (Fig. 2d) and with a threefold increase in density (Extended Data Fig. 1c). The densified wood has a unique microstructure: the fully collapsed wood cell walls are tightly intertwined along their cross-section (Fig. 2e and Extended Data Fig. 1g, j) and densely packed along their length direction (Fig. 2f and Extended Data Fig. 1h, i). By contrast, pure hot-pressing of natural wood without partial lignin/hemicellulose removal can only modestly densify the wood, leaving many gaps in between collapsed cell walls (Extended Data Fig. 2a–c). Wide-angle X-ray diffraction (Fig. 2g), small-angle X-ray scattering and scanning electron microscopy (SEM) (Fig. 2i and Extended Data Fig. 1k, l) further reveal that, at a finer scale, the cellulose nanofibres within the densified wood

remain highly aligned, similar to natural wood but much more densely packed.

The mechanical properties of the densified wood are not only remarkably superior to those of natural wood, but also exceed those of many widely used structural materials (for example, plastics, steel and alloys). Figure 3a compares the tensile stress–strain curves for natural wood and densified wood. Both curves show a linear deformation behaviour before tensile failure. The densified wood demonstrates a record high tensile strength of 587 MPa, which is 11.5 times higher than that of the untreated natural wood (Fig. 3a, b), and also much higher than that of typical plastics<sup>22–24</sup> (such as nylon 6, polycarbonate, polystyrene and epoxy; Fig. 3c) and other densified woods (Extended Data Fig. 6m). A long-standing challenge in engineering material design is the conflict between strength and toughness, because these properties are in general mutually exclusive<sup>25,26</sup>. Interestingly, the large increase in tensile strength of the densified wood is not accompanied by a decrease in toughness. Both the work of fracture and the elastic stiffness of the densified wood are more than ten times higher than those of natural wood (Fig. 3b and Extended Data Fig. 3a). Charpy impact tests of the densified wood yield an impact toughness of  $11.41 \pm 0.5 \text{ J cm}^{-2}$ , 8.3 times higher than that of the natural wood ( $1.38 \pm 0.3 \text{ J cm}^{-2}$ ) (Extended Data Fig. 3d). The scratch hardness and hardness modulus of the densified wood are 30 times and 13 times higher than those of natural wood, respectively (Extended Data Fig. 3b, c, e). The flexural strength of the densified wood is about 6 times and 18 times higher than that of natural wood along and

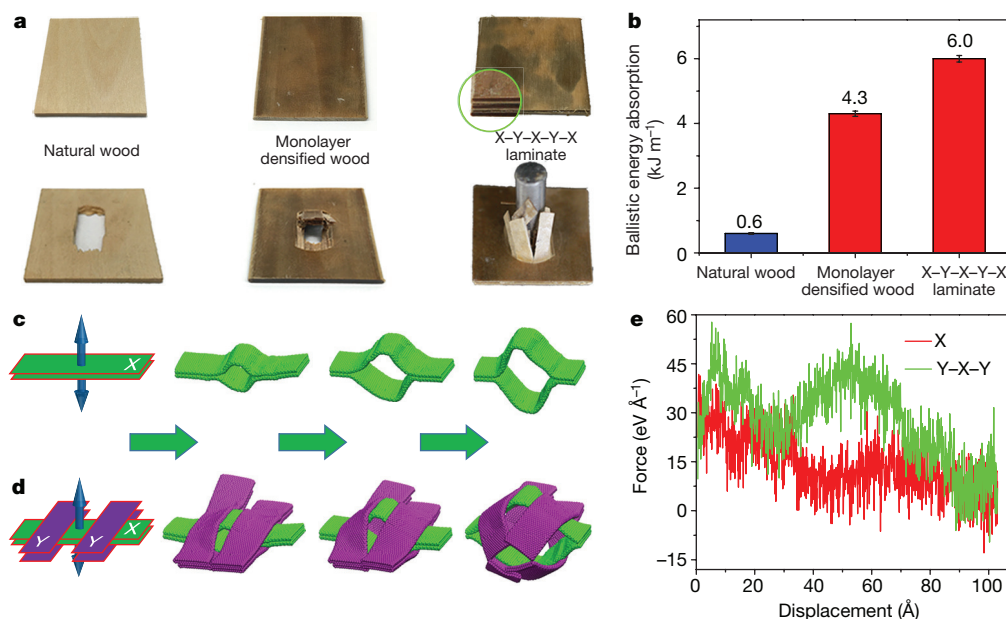


**Figure 3 | Superb mechanical properties of densified wood and mechanistic understanding.** **a**, Tensile stress–strain curves for natural wood and densified wood. **b**, Compared with natural wood (strength,  $46.7 \pm 4.5$  MPa; work of fracture,  $0.39 \pm 0.04$   $\text{MJ m}^{-3}$ ), the densified wood (strength,  $548.8 \pm 47.2$  MPa; work of fracture,  $3.9 \pm 0.2$   $\text{MJ m}^{-3}$ ) has greatly improved strength (12 times) and work of fracture (10 times). **c**, Comparison of the tensile strength of densified wood ( $548.8 \pm 47.2$  MPa) with other widely used polymer-based materials. **d**, **e**, SEM images of the

tensile fracture surface of the natural wood and densified wood samples, respectively (RT plane). **f**, Simulation model of representative deformation and failure process in natural wood, containing a bundle of seven hollow wood lumina. **g**, Corresponding simulation model for densified wood with seven wood lumina fully collapsed. **h**, **i**, Simulated stress–strain curves for the relative sliding in the hollow lumina bundle and in collapsed lumina (**h**), which indicate a 7.5-fold increase in strength and work of fracture as a result of the densification treatment (**i**).

perpendicular to the growth direction, respectively (Extended Data Fig. 3f–n). The compressive strength of the densified wood is about 5.5 times and 33–52 times higher than that of natural wood along and perpendicular to the growth direction, respectively (Extended Data Fig. 4). We found that partial lignin removal allows for the highest density of the resulting wood with the best tensile strength, work of fracture and axial compressive strength (Extended Data Figs 2f–h and 4j). Without lignin removal, it is difficult to hot-press natural wood into a completely compact wood (Extended Data Fig. 2a–c shows numerous voids left between the cell walls). However, total lignin removal leads to wood that can be easily crushed during hot-pressing, probably owing to the absence of lignin as a binder (Extended Data Fig. 2d, e). The intrinsically light weight of cellulose also results in a specific strength of the densified wood ( $451 \text{ MPa cm}^{-3} \text{ g}^{-1}$ ) even higher than that of lightweight titanium alloy (about  $244 \text{ MPa cm}^{-3} \text{ g}^{-1}$ ) (Fig. 1b)<sup>27–30</sup>. The densified wood is stable under moisture attack. For example, subjected to 95% relative humidity (RH) for 128 h, the densified wood swells to produce an increase of only 8.4% in thickness, with only a modest drop in tensile strength ( $493.1 \pm 20.3$  MPa, still 10.6 times higher than that of natural wood in ambient environment). Furthermore, by applying a standard surface treatment (painting), the densified wood is shown to be immune from moisture attack in the accelerated tests (Extended Data Fig. 5). More comprehensive studies demonstrate that our top-down two-step processing approach is universally effective for various species of wood (both hardwood and softwood) and can greatly enhance their strength and toughness simultaneously (Extended Data Fig. 6a–l).

A comparison between the tensile fracture surface of natural wood and that of the densified wood offers insights into the strengthening and toughening mechanisms in the densified wood. Tensile failure of natural wood initiates from relative sliding among open wood lumina followed by the pulling out and tearing of the wood lumina along the fracture surface (Fig. 3d and Extended Data Fig. 7a, b), while the tensile failure of the densified wood results from relative sliding among densely packed wood cell walls followed by the pulling out and fracture of the cell walls along the fracture surface (Fig. 3e and Extended Data Fig. 7c, d). Given that cellulose is the dominant constituent of the densified wood, the corresponding toughening and strengthening mechanisms can be understood as follows. The densely packed and intertwined wood cell walls in the densified wood at the microscale lead to a high degree of alignment of cellulose nanofibres and thus drastically increase the interfacial area among nanofibres. At the molecular scale, owing to the rich hydroxyl groups in cellulose molecular chains, relative sliding among densely packed wood cell walls involves an enormous number of repeating events of hydrogen-bond formation, breaking and reformation at the molecular scale<sup>8</sup> (Fig. 1a and Extended Data Fig. 1m). Consequently, the total energy needed to fracture the densified wood is much higher than that needed to fracture natural wood. In other words, the densified wood is much tougher than natural wood. The densely packed microstructure also greatly reduces both the quantity and size of defects (ranging from vessels to tracheids and pits on cell walls; Extended Data Fig. 1d–i) in the densified wood, producing a much higher strength than that of natural wood. Further modelling of the mechanics of the envisioned deformation and failure processes



**Figure 4 | Ballistic test.** **a**, Photographs of the natural wood, monolayer densified wood and the X-Y-X-Y-X laminated densified wood before (top row) and after (bottom row) the ballistic test. **b**, Comparison of ballistic energy absorption of three types of wood (natural wood,  $0.60 \pm 0.03 \text{ kJ m}^{-2}$ ; monolayer densified wood,  $4.30 \pm 0.08 \text{ kJ m}^{-2}$ ; X-Y-X-Y-X laminate,  $6.0 \pm 0.1 \text{ kJ m}^{-2}$ ). **c**, **d**, Mechanistic understanding of enhanced ballistic resistance in laminated densified wood. Simulation trajectory snapshots during the separation of two neighbouring collapsed

wood cell walls in parallel are shown (**c**) along with simulation snapshots for the same two wood cell walls sandwiched between another two pairs of collapsed wood cell walls along the perpendicular direction at different separation displacements (**d**). **e**, The separation force as a function of separation displacement. The area below the curve indicates energy dissipation. The sandwiched configuration (corresponding to laminated densified wood) could dissipate more energy than the parallel configuration (corresponding to monolayer densified wood).

in natural wood and densified wood (details in Methods) quantitatively verifies the above strengthening and toughening mechanisms. It is shown that both the maximum nominal stress (indicating strength) and the energy dissipation (indicating toughness) associated with sliding between the densely packed collapsed wood lumina are about 7.5 times higher than those associated with hollow wood lumina (Fig. 3f–i and Extended Data Fig. 8a–c). Hydrogen bonds formed between neighbouring cellulose nanofibres make a pivotal contribution to the remarkably enhanced strength and toughness (Extended Data Fig. 8d–f).

The well-aligned cellulose nanofibres dictate the anisotropic mechanical properties of densified wood (Extended Data Fig. 9a–c). To explore the full potential of the exceptional mechanical properties of densified wood, we laminated two layers of natural wood with perpendicular wood fibre orientations, and followed the same processing process to obtain a bilayer densified wood (referred to as X-Y). Tensile strengths of the X-Y densified wood along two perpendicular wood fibre directions are shown to be nearly the same ( $221.6 \pm 20.0 \text{ MPa}$  and  $225.6 \pm 18.0 \text{ MPa}$ , respectively, Extended Data Fig. 9d–f), and much higher than the T-direction strength of monolayer densified wood ( $43.3 \pm 2.0 \text{ MPa}$ ) or that of natural wood ( $5.1 \pm 0.4 \text{ MPa}$ ).

These strong and tough yet lightweight densified woods hold promise as materials for low-cost armour and ballistic energy absorption. To demonstrate such a potential, we used the same processing approach to make a five-layer densified wood with fibre orientation alternating by  $90^\circ$  from layer to layer (referred to as X-Y-X-Y-X). We performed ballistic tests (see Methods) on natural wood, monolayer densified wood and X-Y-X-Y-X densified wood in an air-gun ballistic tester (Extended Data Fig. 10a and Fig. 4a). The ballistic energy absorption per unit sample thickness for monolayer densified wood is  $4.3 \pm 0.08 \text{ kJ m}^{-2}$ , a remarkable sevenfold increase from that of natural wood ( $0.6 \pm 0.03 \text{ kJ m}^{-2}$ ). High-speed-camera videos of the ballistic tests (see Supplementary Video 1) and further characterization of the fractured samples (Extended Data Fig. 10b–g) reveal that in the monolayer densified wood, the perforation opening

by the steel projectile is smaller than that in the natural wood, and the wood surface is severely chapped, indicating much stronger bonding between highly packed wood cell walls (Fig. 4a). The ballistic resistance of the X-Y-X-Y-X densified wood is shown to be even higher and also more isotropic (Extended Data Fig. 10h). The projectile can break through the sample surface but is eventually trapped inside the sample without complete perforation. The resulting ballistic energy absorption is  $6.0 \pm 0.1 \text{ kJ m}^{-2}$ , ten times higher than that of natural wood (Fig. 4b). Further mechanics modelling attributes this enhanced and isotropic ballistic resistance to the reinforcement effect acting between the neighbouring wood layers of alternating orientation (Fig. 4c–e and Extended Data Fig. 10i–l).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 May; accepted 8 December 2017.**

- Wang, J., Cheng, Q. & Tang, Z. Layered nanocomposites inspired by the structure and mechanical properties of nacre. *Chem. Soc. Rev.* **41**, 1111–1129 (2012).
- Millitzer, M. A synchrotron look at steel. *Science* **298**, 975–976 (2002).
- Podsiadlo, P. et al. Ultrastrong and stiff layered polymer nanocomposites. *Science* **318**, 80–83 (2007).
- Zhu, Y. T. & Liao, X. Nanostructured metals: retaining ductility. *Nat. Mater.* **3**, 351–352 (2004).
- Fratzl, P. & Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **52**, 1263–1334 (2007).
- Gao, H., Ji, B., Jäger, I. L., Arzt, E. & Fratzl, P. Materials become insensitive to flaws at nanoscale: lessons from nature. *Proc. Natl Acad. Sci. USA* **100**, 5597–5600 (2003).
- Meyers, M. A., McKittrick, J. & Chen, P.-Y. Structural biological materials: critical mechanics-materials connections. *Science* **339**, 773–779 (2013).
- Zhu, H. et al. Anomalous scaling law of strength and toughness of cellulose nanopaper. *Proc. Natl Acad. Sci. USA* **112**, 8971–8976 (2015).
- Hon, D. N.-S. & Shiraishi, N. *Wood and Cellulosic Chemistry* (CRC Press, 2000).
- Erickson, E. *Mechanical Properties of Laminated Modified Wood* (US Department of Agriculture, Forest Service, Forest Products Laboratory, 1965).
- Fang, C.-H., Mariotti, N., Cloutier, A., Koubaa, A. & Blanchet, P. Densification of wood veneers by compression combined with heat and steam. *Eur. J. Wood Wood Prod.* **70**, 155–163 (2012).

12. Bekhta, P., Hiziroglu, S. & Shepelyuk, O. Properties of plywood manufactured from compressed veneer as building material. *Mater. Des.* **30**, 947–953 (2009).
13. Kultikova, E. V. *Structure and Properties Relationships of Densified Wood* (Virginia Polytechnic Institute and State University, 1999).
14. Paril, P. *et al.* Comparison of selected physical and mechanical properties of densified beech wood plasticized by ammonia and saturated steam. *Eur. J. Wood Wood Prod.* **72**, 583–591 (2014).
15. Navi, P. & Heger, F. Combined densification and thermo-hydro-mechanical processing of wood. *MRS Bull.* **29**, 332–336 (2004).
16. Gong, M., Lamason, C. & Li, L. Interactive effect of surface densification and post-heat-treatment on aspen wood. *J. Mater. Process. Technol.* **210**, 293–296 (2010).
17. O'Connor, J. P. *Improving wood strength and stiffness through viscoelastic thermal compression*. Masters thesis, Oregon State Univ., [https://ir.library.oregonstate.edu/concern/graduate\\_thesis\\_or\\_dissertations/8623j074v](https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/8623j074v) (2007).
18. Laine, K., Segerholm, K., Wälinder, M., Rautkari, L. & Hughes, M. Wood densification and thermal modification: hardness, set-recovery and micromorphology. *Wood Sci. Technol.* **50**, 883–894 (2016).
19. Kutnar, A. & Kamke, F. A. Compression of wood under saturated steam, superheated steam, and transient conditions at 150 °C, 160 °C, and 170 °C. *Wood Sci. Technol.* **46**, 73–88 (2012).
20. Hill, C. A. *et al.* The water vapour sorption properties of thermally modified and densified wood. *J. Mater. Sci.* **47**, 3191–3197 (2012).
21. Laine, K. *et al.* Measuring the thickness swelling and set-recovery of densified and thermally modified Scots pine solid wood. *J. Mater. Sci.* **48**, 8530–8538 (2013).
22. Li, Y. & Shimizu, H. Improvement in toughness of poly(L-lactide) (PLLA) through reactive blending with acrylonitrile–butadiene–styrene copolymer (ABS): morphology and properties. *Eur. Polym. J.* **45**, 738–746 (2009).
23. Kim, K. H. & Jo, W. H. A strategy for enhancement of mechanical and electrical properties of polycarbonate/multi-walled carbon nanotube composites. *Carbon* **47**, 1126–1134 (2009).
24. Ku, H., Wang, H., Pattarachaiyakoo, N. & Trada, M. A review on the tensile properties of natural fiber reinforced polymer composites. *Composites B* **42**, 856–873 (2011).
25. Ritchie, R. O. The conflicts between strength and toughness. *Nat. Mater.* **10**, 817–822 (2011).
26. Wang, Y., Chen, M., Zhou, F. & Ma, E. High tensile ductility in a nanostructured metal. *Nature* **419**, 912–915 (2002).
27. Dursun, T. & Soutis, C. Recent developments in advanced aircraft aluminium alloys. *Mater. Des.* **56**, 862–871 (2014).
28. Frommeyer, G. & Brück, U. Microstructures and mechanical properties of high-strength Fe–Mn–Al–C light-weight TRIPLEX steels. *Steel Res. Int.* **77**, 627–633 (2006).
29. Gil, F., Manero, J., Ginebra, M. & Planell, J. The effect of cooling rate on the cyclic deformation of beta-annealed Ti–6Al–4V. *Mater. Sci. Eng. A* **349**, 150–155 (2003).
30. Kim, S.-H., Kim, H. & Kim, N. J. Brittle intermetallic compound makes ultrastrong low-density steel with large ductility. *Nature* **518**, 77–79 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank R. Briber for suggestions and R. J. Bonenberger for help with mechanical tests. We acknowledge the support of the Maryland NanoCenter and its AIMLab. J.S. acknowledges financial support from the China Scholarship Council.

**Author Contributions** J.S., C.C., S.Z. and M.Z. contributed equally to this work. L.H., J.S., C.C. and M.Z. contributed to the initiating idea. J.S. and C.C. contributed to the wood densification and mechanical measurements. Yo.L., U.R., Z.J., N.Q., U.H.L., H.A.B. and T.L. contributed to the mechanical tensile and ballistic tests. J.D. and Y.K. contributed to the 3D illustrations. Yi.L., C.C., Y.Y. and A.G. contributed to characterization via SEM. J.Y.Z. performed the compositional analysis. A.V. and A.M. contributed to the indentation and scratch hardness tests. S.Z. and T.L. contributed to both mechanical simulations and analysis. H.L. and M.L.M. contributed to XRD measurement and analysis. T.L., L.H., J.S. and C.C. contributed to the writing of the paper. All authors contributed to commenting on the final manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.H. ([binghu@umd.edu](mailto:binghu@umd.edu)) or T.L. ([lit@umd.edu](mailto:lit@umd.edu)).

**Reviewer Information** *Nature* thanks A. Cloutier, S. Eichhorn and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Materials and chemicals.** Basswood (*Tilia*), oak (*Quercus*), poplar (*Populus*), western red cedar (*Thuja plicata*) and eastern white pine (*Pinus strobus*) were used for the fabrication of densified wood. Sodium hydroxide (>97%, Sigma-Aldrich) and sodium sulfite (>98%, Sigma-Aldrich) and deionized (DI) water were used for processing the wood.

**Two-step process towards densified wood.** First, natural wood blocks (typical sample dimension: 120.0 mm by 44.0 mm by 44.0 mm) were immersed in a boiling aqueous solution of mixed 2.5 M NaOH and 0.4 M Na<sub>2</sub>SO<sub>3</sub> for 7 h, followed by immersion in boiling deionized water several times to remove the chemicals. Next, the wood blocks were pressed at 100 °C under a pressure of about 5 MPa for about 1 day to obtain the densified wood (115.6 mm by 46.5 mm by 9.5 mm). By adjusting the boiling times, densified wood with different degrees of lignin removal can be obtained.

**Mechanical tests.** The tensile, bending and compressive properties of the wood samples were measured using a Tinius Olsen H5KT tester. The dimensions for tensile samples were approximately 100 mm by 6 mm by 1.5 mm. The samples were clamped at both ends and stretched along the sample length direction until they fractured with a constant test speed of 5 mm min<sup>-1</sup> at room temperature. The dimensions for bending samples were approximately 35 mm by 5 mm by 4 mm. Three-point bending tests were conducted for these samples, with the span between the two bottom rollers 20 mm and the top roller pressing down at the centre at a speed of 1 mm min<sup>-1</sup>. The flexural stress is defined as the maximum tensile stress at the bottom surface of the sample right below the top roller. The dimensions for compressive samples were approximately 9 mm long, 9 mm wide and 4.5 mm thick, and the samples were compressed along the thickness direction at a speed of 1 mm min<sup>-1</sup>.

**Scratch hardness test.** The scratch resistance of wood samples was evaluated according to the Standard Test Method for Scratch Hardness, ASTM G171-03(2009) using a linear reciprocating tribometer (Rtec Instruments Multi-Function Tribometer). The test was performed by applying a normal load on a diamond sphero-conical tip indenter and moving the wood surface laterally relative to the indenter at a constant speed. The width of the scratch was then measured using an optical microscope and the scratch hardness number (in gigapascals) was calculated as  $kP/w^2$ , where  $P$  is the applied normal force,  $w$  is the scratch width and  $k$  is the geometric constant. Each scratch hardness value was determined as an arithmetic mean of a set of three scratches made side by side at different locations. The lateral speed of the sample and the stroke length of the scratch were chosen as 0.2 mm s<sup>-1</sup> and 7 mm, respectively.

**Hardness modulus test.** The hardness modulus was measured using a modified version of the standard procedure described in the ASTM D1037-121 with an Rtec Instruments Multi-Function Tribometer. The standard recommends indenting specimens of thickness 3–6 mm, using a ball of diameter 1.3 mm, to a depth of 2.5 mm. Since our test specimens are 5 mm thick, we used a smaller ball of diameter 4.76 mm with a penetration depth of 1.05 mm, which corresponds to an average Hertzian contact pressure equal to that in the standard test. The rate of penetration was constant at the recommended value of 1.3 mm min<sup>-1</sup>. The penetration force versus depth was plotted and the slope of the linear portion of this curve was calculated as the hardness modulus (pounds per inch). Five indentations were made on each specimen and the average value was reported.

**Charpy impact test.** The Charpy impact test of the wood samples was performed on a Tinius Olsen pendulum impact tester. The dimensions of the samples were 14 mm × 4.5 mm × 100 mm.

**Ballistic test.** We performed the ballistic tests on wood samples using a gas gun, which comprises a pressure indicator frame, two cylinders filled with compressed nitrogen (N<sub>2</sub>), a pressure chamber 127 mm in diameter and barrel-length 190.5 mm, a nozzle of length 1,156 mm and internal diameter 12.5 mm and a holder specifically designed to clamp the sample. The pressure indicator frame has dials with which we adjusted the pressure inside the two N<sub>2</sub> cylinders. The left cylinder is used to pressurize the volume inside the barrel chamber when the projectile is fired and the right N<sub>2</sub> cylinder controls the pressure for the firing valve, helping it

open instantaneously when fired. The chamber pressure was set to about 2.22 MPa. Once opened, the valve releases the pressure and accelerates the projectile. We used projectiles cylindrical in shape, made of stainless steel, with a diameter of 11.85 mm, length 51.77 mm and mass 0.046 kg. The dimensions of the samples were approximately 44 mm by 44 mm by 3 mm. The whole ballistic process was captured using two Phantom v12 cameras. The Phantom Camera Control software (<http://www.phantomhighspeed.com/products/accessories-and-options/camera-control-software>), developed for such high-speed digital cameras, captured the velocities of the projectile before and after perforating the sample. The ballistic energy absorption of the test sample is defined by the kinetic energy loss after a cylindrical steel projectile perforates the sample. The equation for the ballistic energy absorption normalized by sample thickness is  $m(v_1^2 - v_2^2)/2t$ , where  $m$  is the mass of the projectile,  $t$  is the thickness of the sample and  $v_1$  and  $v_2$  are the velocities of the projectile before and after perforating the sample, respectively.

**Accelerated test against moisture.** The humidity chamber (LHS-150HC-II) was set up at 20 °C, 95% RH. Then pre-cut samples with dimensions of approximately 15 mm by 15 mm by 4.6 mm were placed in the humidity chamber. The dimensions and weight of the samples after various intervals in the humidity chamber were recorded. Following a painting method widely used in the wood industry, we coated the densified wood with a thin layer of oil-based paint (Polyurethane, Minwax). After the paint was totally dry, the sample was put into the humidity chamber and measured at regular intervals.

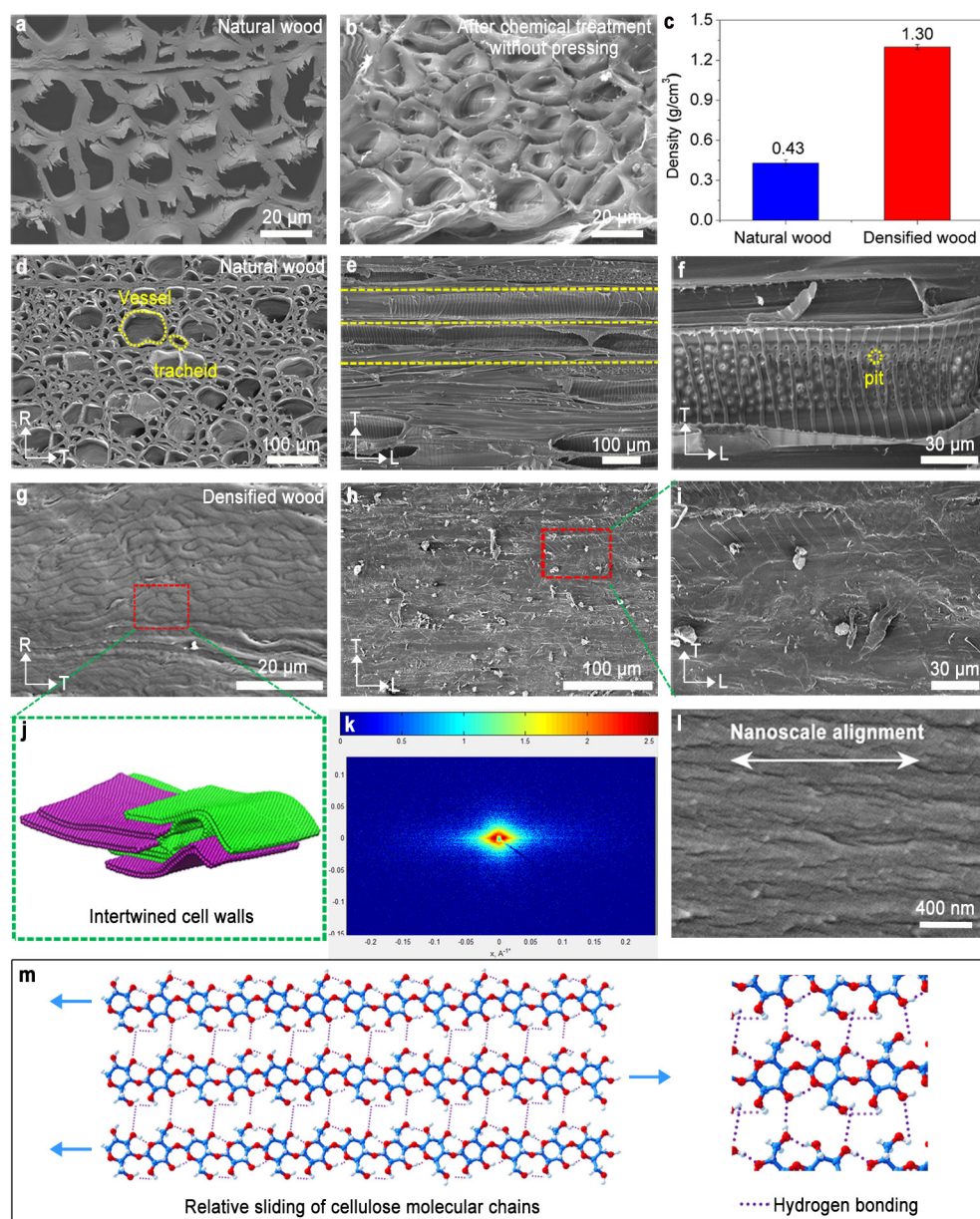
**Characterizations.** A scanning electron microscope (SEM, Hitachi SU-70) was used to characterize the morphologies of the wood samples. Small-angle X-ray scattering (SAXS) patterns for three samples of each wood were collected using Rigaku MicroMax 007HF (operating voltage at 40 kV, current at 30 mA, CuKα,  $\lambda = 0.1541$  nm). The angle between the incident X-ray beam and the width direction on the sample was kept at 90°. The raw azimuthal intensity distribution was extracted and the baseline is subtracted. Wide-angle X-ray diffraction patterns were collected on multi-filament bundles using a Rigaku RAPID II (operating voltage at 40 kV, current at 30 mA, CuKα,  $\lambda = 0.1541$  nm) equipped with a curved detector manufactured by Rigaku Americas Corporation. Compositional analysis of natural wood and chemical-treated wood was carried out on a high-performance liquid chromatography (HPLC) system (Ultimate 3000, Thermo Scientific, USA).

**Mechanics modelling.** We used a generic coarse-grained simulation scheme to qualitatively reveal the underlying mechanism for the enhancement in mechanical properties. The wood fibre is modelled as a tube made of coarse-grained beads that assume a hexagonal lattice structure (Extended Data Fig. 8d). The bonded energy terms of the coarse-grained scheme consist of a two-body bond energy and three-body angle energy and a four-body torsion energy as follows:

$$U_{\text{bonded}}(r_{ij}, \theta_{ijk}) = \sum \frac{1}{2} K_{\text{bond}} (r_{ij} - r_0)^2 + \sum \frac{1}{2} K_{\theta} (\cos \theta_{ijk} - \cos \theta_0)^2 + \sum_{n=1,5} A_n \cos^{n-1} \Phi$$

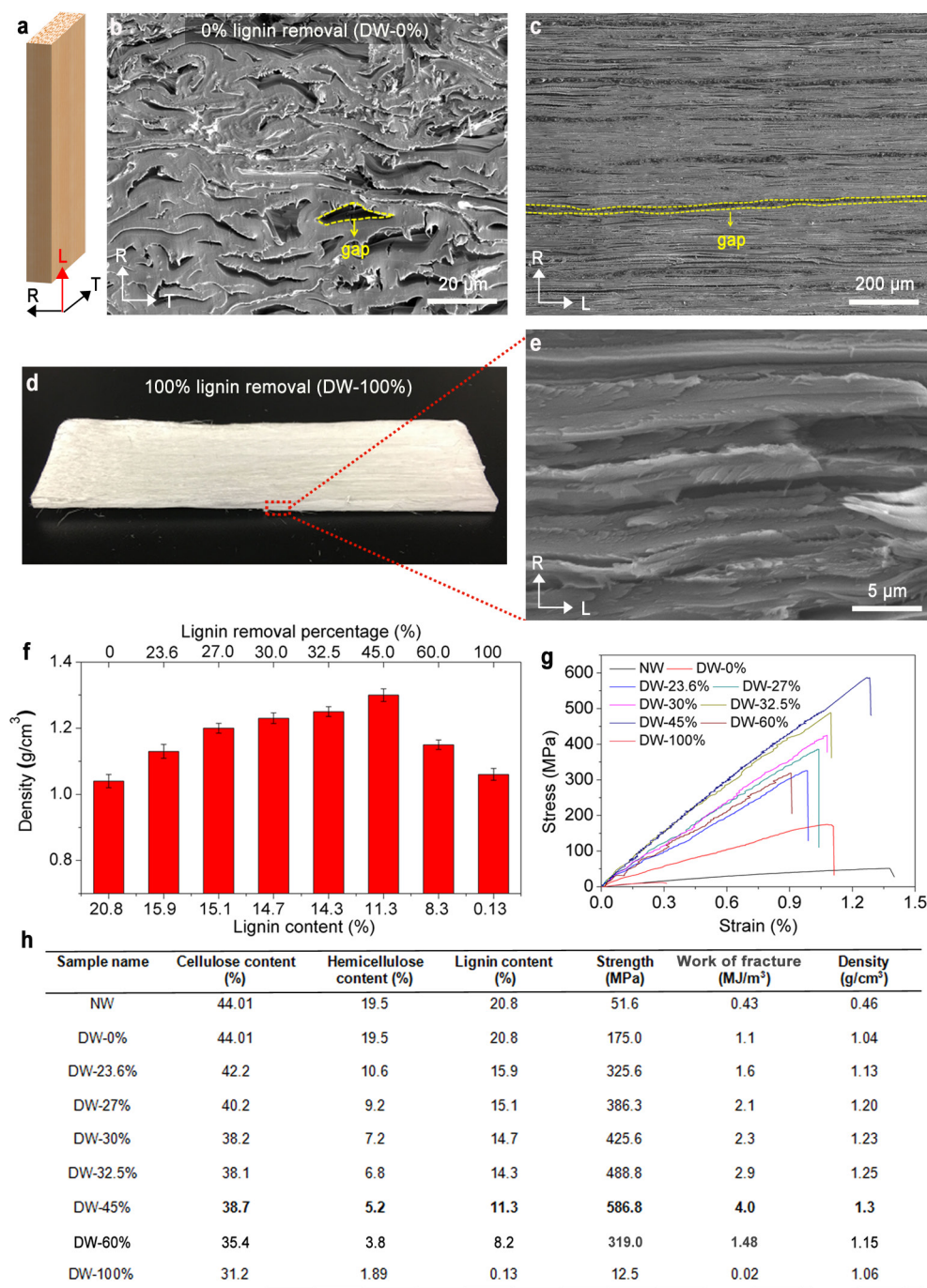
where  $K_{\text{bond}}$  and  $K_{\theta}$  are the bond force constant and the angle force constant, respectively,  $r_{ij}$  is the distance between the  $i$ th and  $j$ th coarse-grained beads while  $r_0$  is the corresponding equilibrium value of  $r_{ij}$ ;  $\theta_{ijk}$  is the angle formed between the  $i$ - $j$  bond and the  $j$ - $k$  bond while  $\theta_0$  is the corresponding equilibrium value of  $\theta_{ijk}$  (which is 120° for all cases).  $A_n$  are coefficients ( $n = 1, 2, 3, 4, 5$ ) for the dihedral angle  $\Phi$ . The non-bonded term includes the long-range van der Waals Lennard-Jones-type interaction  $4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$  between coarse-grained beads (cut-off distance 1 nm,  $\epsilon$  denotes the interaction strength and  $\sigma$  denotes the distance where the interaction energy crosses zero) and a short-range (cut-off distance 0.24 nm) Morse-type potential  $D_0[e^{-2\alpha(r-r_{\text{Morse}})} - 2e^{-\alpha(r-r_{\text{Morse}})}]$ , which is used to model the hydrogen-bond interaction among wood fibres. The simulation is done at 300 K by canonical ensemble and by Nose-Hoover thermostat. Extended Data Fig. 8g lists the values of the coarse-grained parameters used in the simulations.

**Data availability.** The data that support the findings of this study are available from the corresponding authors on reasonable request.



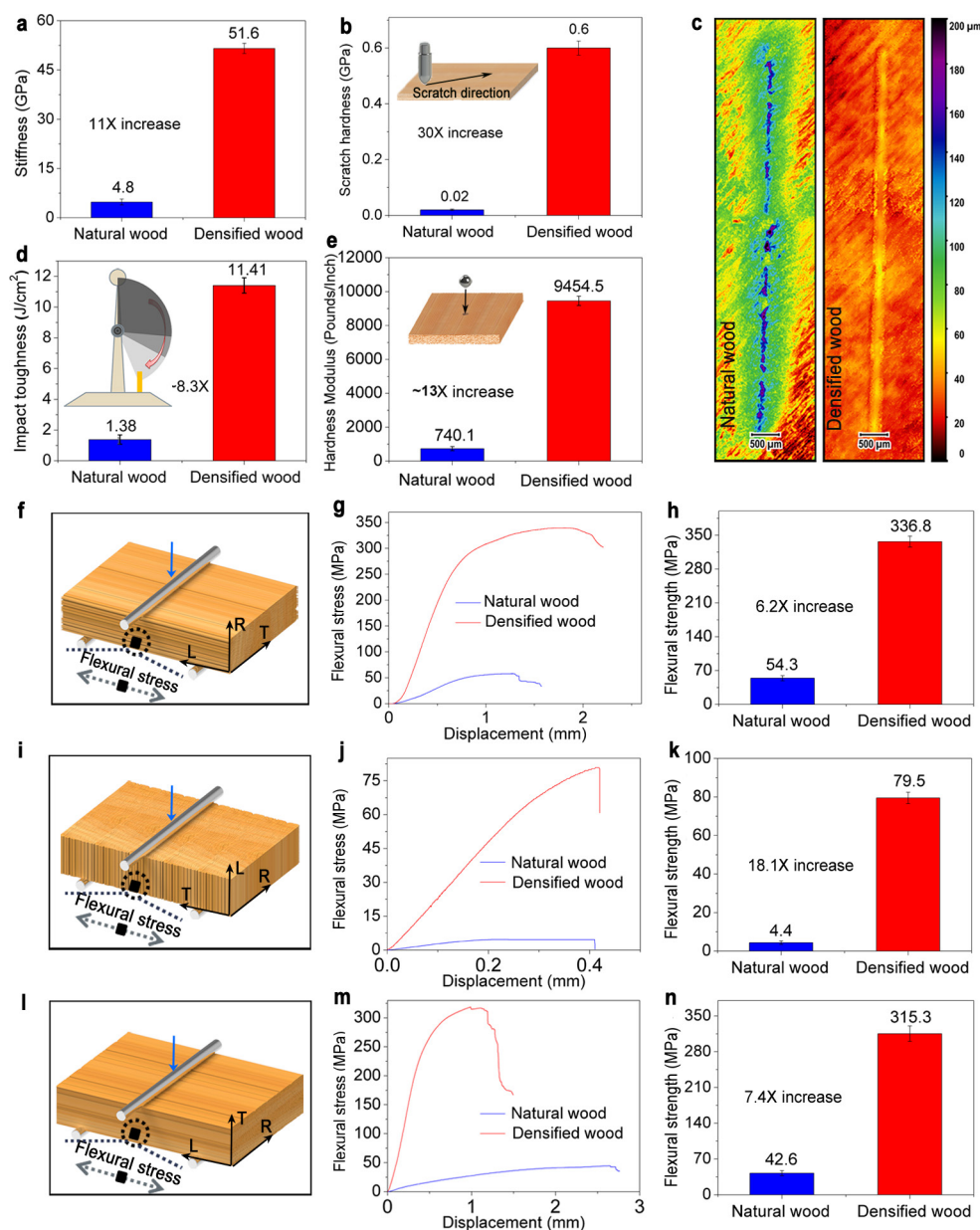
**Extended Data Figure 1 | Structural characterization of natural wood and densified wood.** **a, b**, Comparison of SEM images of natural wood (**a**) and the wood after partial lignin removal but without lateral hot-pressing (**b**) shows that lignin between the cell walls is partially removed. **c**, Comparison of densities of natural ( $0.43 \pm 0.02 \text{ g cm}^{-3}$ ) and densified woods ( $1.30 \pm 0.02 \text{ g cm}^{-3}$ ). **d–f**, SEM images of the cross-section of natural wood in the RT (**d**) and TL (**e, f**) planes show intrinsic defects such as vessels and tracheids along the L direction and pits in the cell walls. **g–j**, The corresponding SEM images of densified wood show that the

hollow lumina are completely collapsed to form highly intertwined wood cell walls (**g**), as verified by the simulation model (**j**), and even the tiny pits in the wood cell walls are eliminated owing to the densification (**h, i**). **k–m**, The small-angle X-ray scattering pattern (**k**) and the high-magnification SEM image (**l**) show well-aligned cellulose nanofibres in densified wood, which greatly facilitate the formation of hydrogen bonds in neighbouring cellulose molecular chains during their relative sliding (**m**).



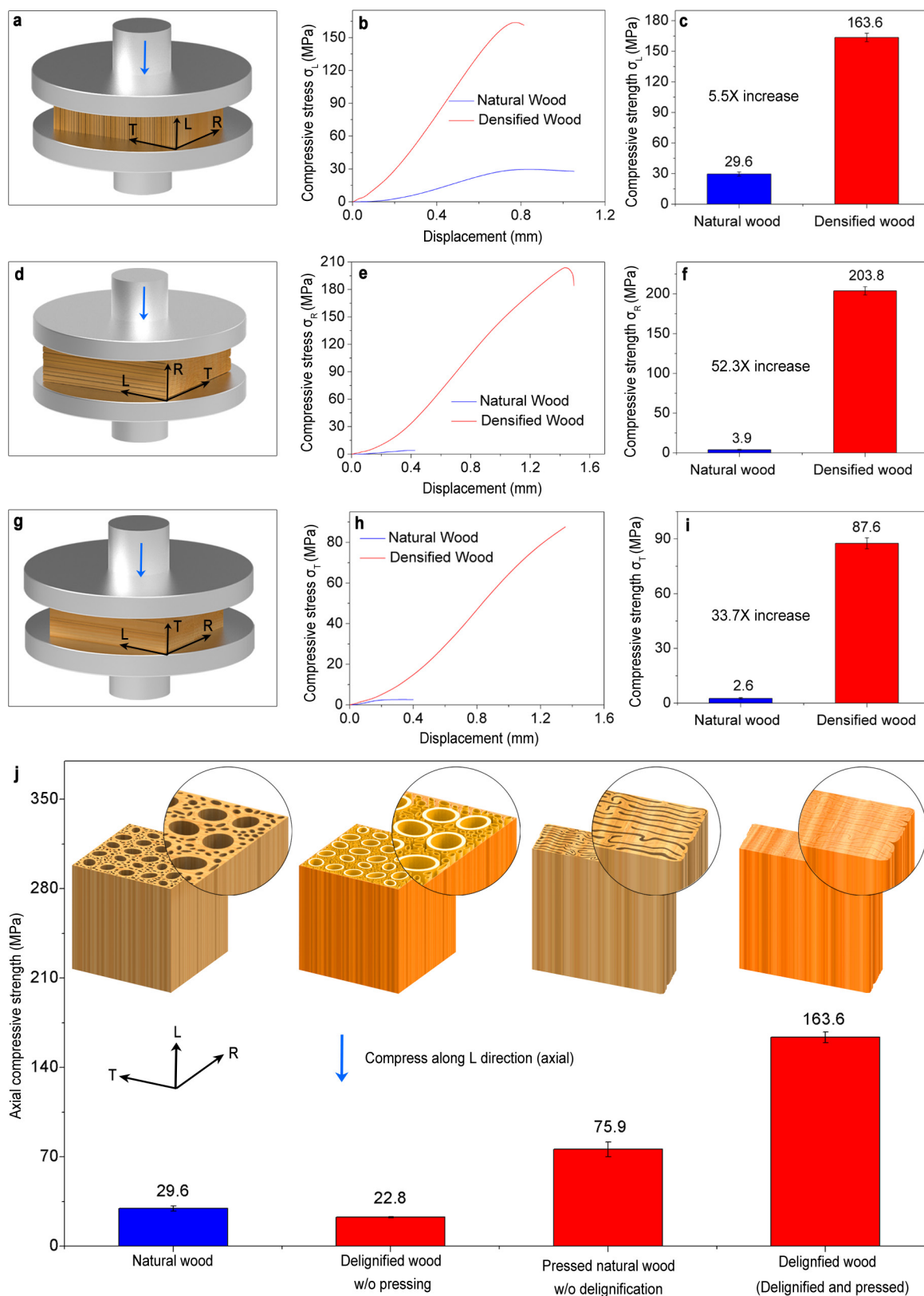
**Extended Data Figure 2 | Effect of degree of lignin removal on wood structure and mechanical properties.** **a**, Schematics of wood sample with the L direction as the tree-growth direction. **b**, **c**, SEM images of the cross-sections in the RT plane (**b**) and the RL plane (**c**) of a pressed wood sample with 0% lignin removal, which show a large number of gaps remaining in between partially collapsed cell walls. **d**, **e**, Photo and SEM image of the densified wood with 100% lignin removal show that the pressed cell walls are separated from each other owing to the absence of lignin as binding

agent. **f**, **g**, Densities (**f**) and tensile stress–strain curves (**g**) of densified woods with various degrees of lignin removal. **h**, Summary of cellulose/hemicellulose/lignin contents as well as strength, work of fracture and density under various degrees of lignin removal. Densified wood with 45% lignin removal is shown to have the highest strength, work of fracture and density. DW- $x$  refers to densified wood with a certain amount ( $x$ ) of lignin removal and subsequent densification, whereas NW refers to natural wood without lignin removal or densification.



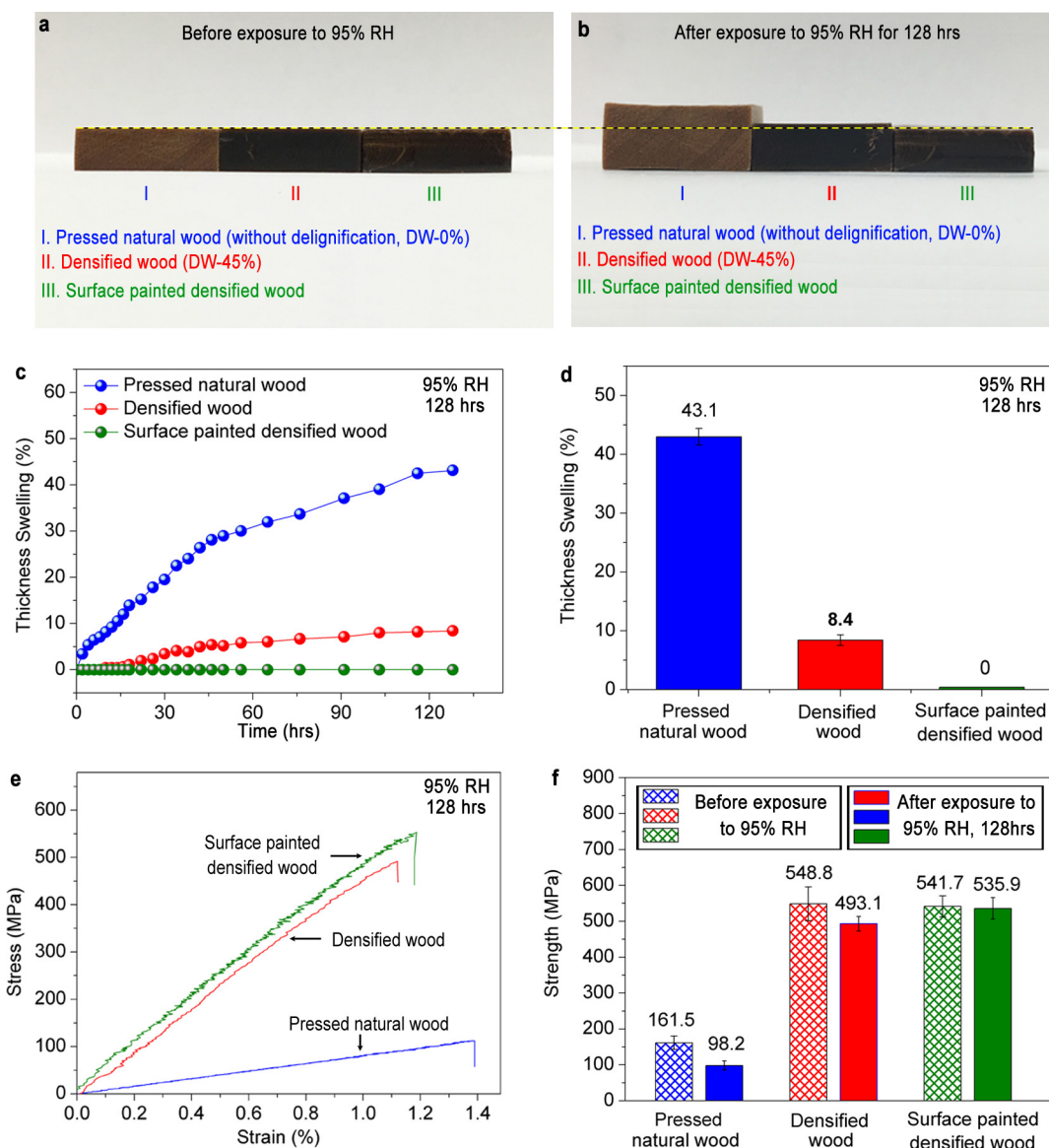
**Extended Data Figure 3 | Comparison of mechanical properties of natural wood and densified wood.** **a**, Stiffness (natural wood,  $4.8 \pm 0.9$  GPa; densified wood,  $51.6 \pm 1.5$  GPa). **b**, Scratch hardness (natural wood,  $0.02 \pm 0.0029$  GPa; densified wood,  $0.6 \pm 0.025$  GPa). **c**, Interferometer images of scratches on natural wood and densified wood, showing the notable decrease of the scratch depth of the densified wood owing to increased hardness. **d**, Charpy impact toughness (densified wood,  $11.41 \pm 0.5$   $\text{J}/\text{cm}^2$ ; natural wood,  $1.38 \pm 0.3$   $\text{J}/\text{cm}^2$ ). **e**, Hardness modulus (natural wood,  $740.1 \pm 115.4$  pounds per inch; densified wood,  $9454.5 \pm 273.3$  pounds per inch). **f**, **i**, **l**, Schematics of bending tests

along three different directions. **g**, **j**, **m**, Corresponding flexural stress as a function of roller displacement (bending deflection) for natural wood and densified wood. **h**, **k**, **n**, Comparison of the corresponding flexural strengths of natural wood (with the roller along the T direction,  $54.3 \pm 5.1$  MPa; perpendicular to wood growth direction,  $4.4 \pm 0.9$  MPa; with the roller along the R direction,  $42.6 \pm 4.9$  MPa; eight samples tested for each direction) and densified wood (with the roller along the T direction,  $336.8 \pm 11.3$  MPa; perpendicular to wood growth direction,  $79.5 \pm 3.0$  MPa; with the roller along the R direction,  $315.3 \pm 14.8$  MPa; eight samples tested for each direction).



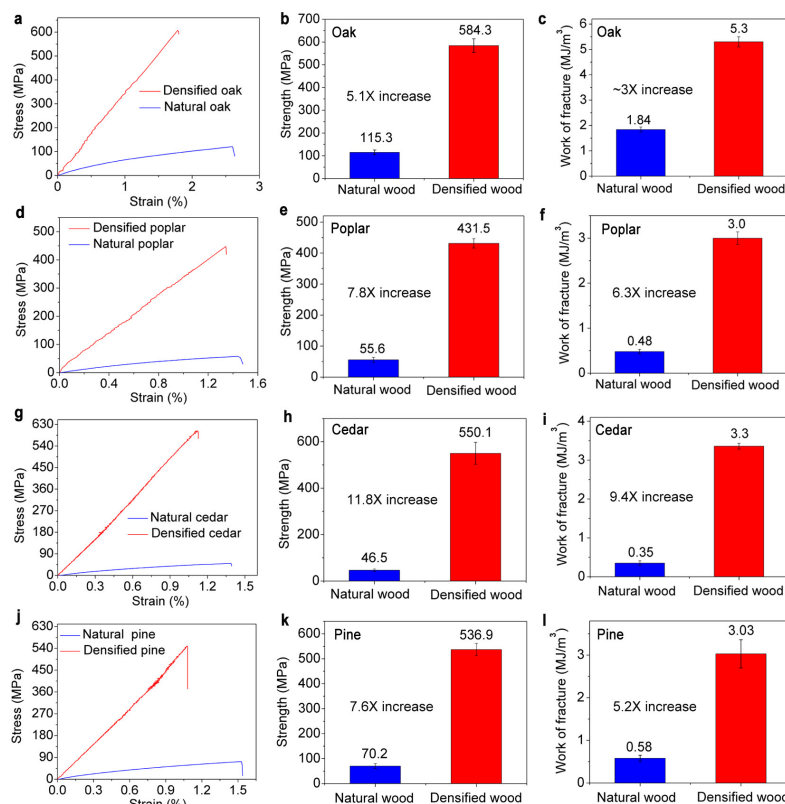
**Extended Data Figure 4 | Compressive strength of natural wood and densified wood.** **a, d, g,** Schematics of compression tests along three different directions. **b, e, h,** Corresponding compressive stress as a function of compressive displacement for natural wood and densified wood. **c, f, i,** Comparison of the corresponding compressive strengths of natural wood (L direction,  $29.6 \pm 2.0$  MPa; R direction,  $3.9 \pm 0.6$  MPa; T direction,  $2.6 \pm 0.4$  MPa; eight samples tested for each

direction) and densified wood (L direction,  $163.6 \pm 4.1$  MPa; R direction,  $203.8 \pm 5.2$  MPa; T direction,  $87.6 \pm 3.0$  MPa; eight samples tested for each direction). **j,** Comparison of axial compressive strengths (along the L direction) of natural wood, delignified wood without hot-pressing, pressed natural wood without delignification, and densified wood (delignified and then hot-pressed). Insets illustrate the representative cross-section features of the four types of wood.



**Extended Data Figure 5 | Dimensional stability and mechanical properties of pressed natural wood, densified wood and surface-painted densified wood against moisture.** **a, b**, Photographs of pressed natural wood without delignification, densified wood (45% lignin removal and then hot-pressed) and surface-painted densified wood before (**a**) and after (**b**) sustaining 95% RH for 128 h. **c**, Change in thickness of the three wood samples over time. **d**, Percentage increase in thickness (pressed natural wood,  $43.1 \pm 1.4\%$ ; densified wood,  $8.4 \pm 0.9\%$ ; surface-painted

densified wood, 0%). **e**, Tensile stress–strain curves of the three wood samples after sustaining 95% RH for 128 h. **f**, Strengths of the three wood samples before (pressed natural wood,  $161.5 \pm 18.8$  MPa; densified wood,  $548.8 \pm 47.2$  MPa; surface-painted densified wood,  $541.7 \pm 29.2$  MPa) and after (pressed natural wood,  $98.2 \pm 12.6$  MPa; densified wood,  $493.1 \pm 20.3$  MPa; surface-painted densified wood,  $535.9 \pm 30.0$  MPa) sustaining 95% RH for 128 h.

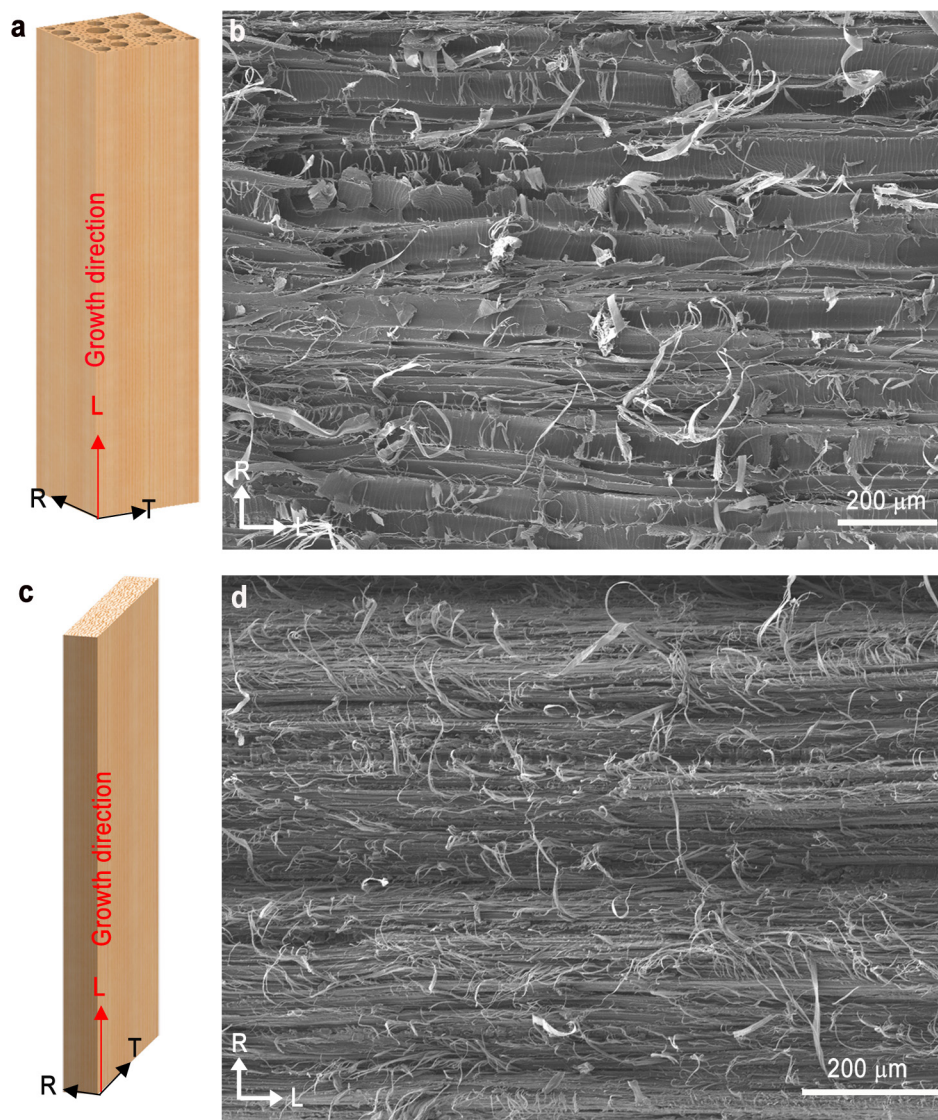


| Treatment   | Lignin Removal | Wood materials     | Ultimate tensile strength (MPa) |                |             | Compression ratio (%) | Structure               | Ref.     |
|---|----------------|--------------------|---------------------------------|----------------|-------------|-----------------------|-------------------------|----------|
|   |                |                    | Before pressing                 | After pressing | Enhancement |                       |                         |          |
| Compressed at ~180 °C and 65% relative humidity             | No             | Sitka spruce       | ~88                             | ~320           | ~3.6        | ~60                   | Densified               | Ref. 10  |
|   | No             | Sweetgum           | ~102                            | ~220           | ~2.2        | ~55                   | Densified               |          |
|   | No             | Yellow birch       | ~151                            | ~307           | ~2          | ~47                   | Densified               |          |
| Fill in resin and compress                                  | No             | Sitka spruce       | ~88                             | ~166           | ~1.9        | <50                   | Densified               | Ref. 10  |
|   | No             | Sweetgum           | ~102                            | ~185           | ~1.8        | <50                   | Densified               |          |
|   | No             | Yellow birch       | ~151                            | ~210           | ~1.4        | <50                   | Densified               |          |
| Combined with heat (160 °C), steam (550 KPa), and pressure. | No             | Aspen              | ~60                             | ~120           | 2           | ~50                   | Densified               | Ref. 11  |
| Cold rolling process  | No             | Birch              | 128.2                           | 130-164        | 1.3         | 5-25                  | Densified               | Ref. 12  |
|   | No             | Alder              | 63.5                            | 64-113.6       | 1.8         | 5-25                  | Densified               |          |
| 90 °C, 95% relative humidity, 93 KPa                        | No             | Loblolly pine      | 117                             | 204            | 1.7         | 50                    | Densified               | Ref. 13  |
|   | No             | Yellow poplar      | 121                             | 197            | 1.6         | 50                    | Densified               |          |
| Ammonia and saturated steam                                 | No             | Beech              | —                               | —              | —           | 15, 30, 50            | Densified with pores    | Ref. 14  |
| Steam & press   | No             | Spruce             | —                               | —              | —           | ~60                   | Densified with pores    | Ref. 15  |
| Densification & post-heat-treatment                         | No             | Aspen              | —                               | —              | —           | 50                    | Densified with pores    | Ref. 16  |
| Viscoelastic Thermal Compression                            | No             | Eastern cottonwood | —                               | —              | —           | ~50                   | Densified with pores    | Ref. 17  |
| Steam & hot-press   | No             | Scots pine         | —                               | —              | —           | 40, 50, 60            | Densified with pores    | Ref. 18  |
| Saturated steam & hot-press                                 | No             | Hybrid poplar      | —                               | —              | —           | ~60                   | Densified with pores    | Ref. 19  |
| Chemical modification & hot-press                           | Yes            | Basswood           | 52                              | 587            | 11.5        | ~80                   | Densified without pores | Our work |
|   | Yes            | Oak                | 122                             | 608            | 5           | ~70                   | Densified without pores | Our work |

**Extended Data Figure 6 | Superb mechanical properties of various species of densified wood.** Comparison of the stress–strain curve, tensile strength and work of fracture for natural and densified woods of various species: the hardwoods oak and poplar, and the softwoods cedar and pine.

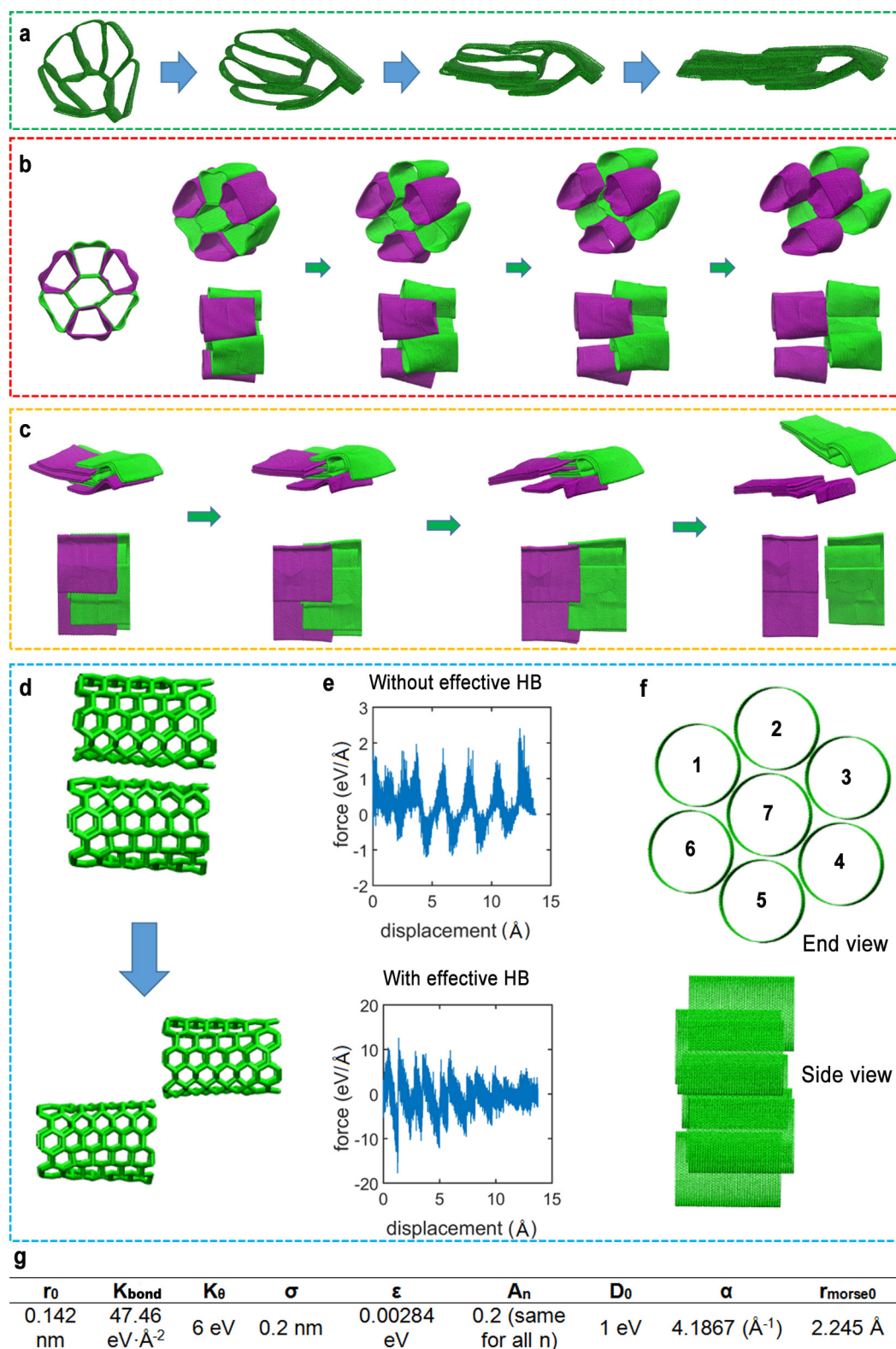
**a–c**, Oak (natural wood strength,  $115.3 \pm 10.2$  MPa; densified wood strength,  $584.3 \pm 29.8$  MPa; natural wood work of fracture,  $1.84 \pm 0.1$  MJ m<sup>-3</sup>; densified wood work of fracture,  $5.3 \pm 0.2$  MJ m<sup>-3</sup>). **d–f**, Poplar (natural wood strength,  $55.6 \pm 8.0$  MPa; densified wood strength,  $431.5 \pm 15.1$  MPa; natural wood work of fracture,  $0.48 \pm 0.05$  MJ m<sup>-3</sup>; densified wood

work of fracture,  $3.0 \pm 0.1$  MJ m<sup>-3</sup>). **g–i**, Cedar (natural wood strength:  $46.5 \pm 5.4$  MPa; densified wood strength:  $550.1 \pm 47.4$  MPa; natural wood work of fracture:  $0.35 \pm 0.06$  MJ m<sup>-3</sup>; densified wood work of fracture,  $3.3 \pm 0.08$  MJ m<sup>-3</sup>). **j–l**, Pine (natural wood strength,  $70.2 \pm 10.0$  MPa; densified wood strength,  $536.9 \pm 24.7$  MPa; natural wood work of fracture,  $0.58 \pm 0.07$  MJ m<sup>-3</sup>; densified wood work of fracture,  $3.03 \pm 0.33$  MJ m<sup>-3</sup>). **m**, Comparison of the structural and mechanical properties of the densified wood in this study and other previously reported<sup>10–19</sup> densified wood materials made from different species of natural wood.



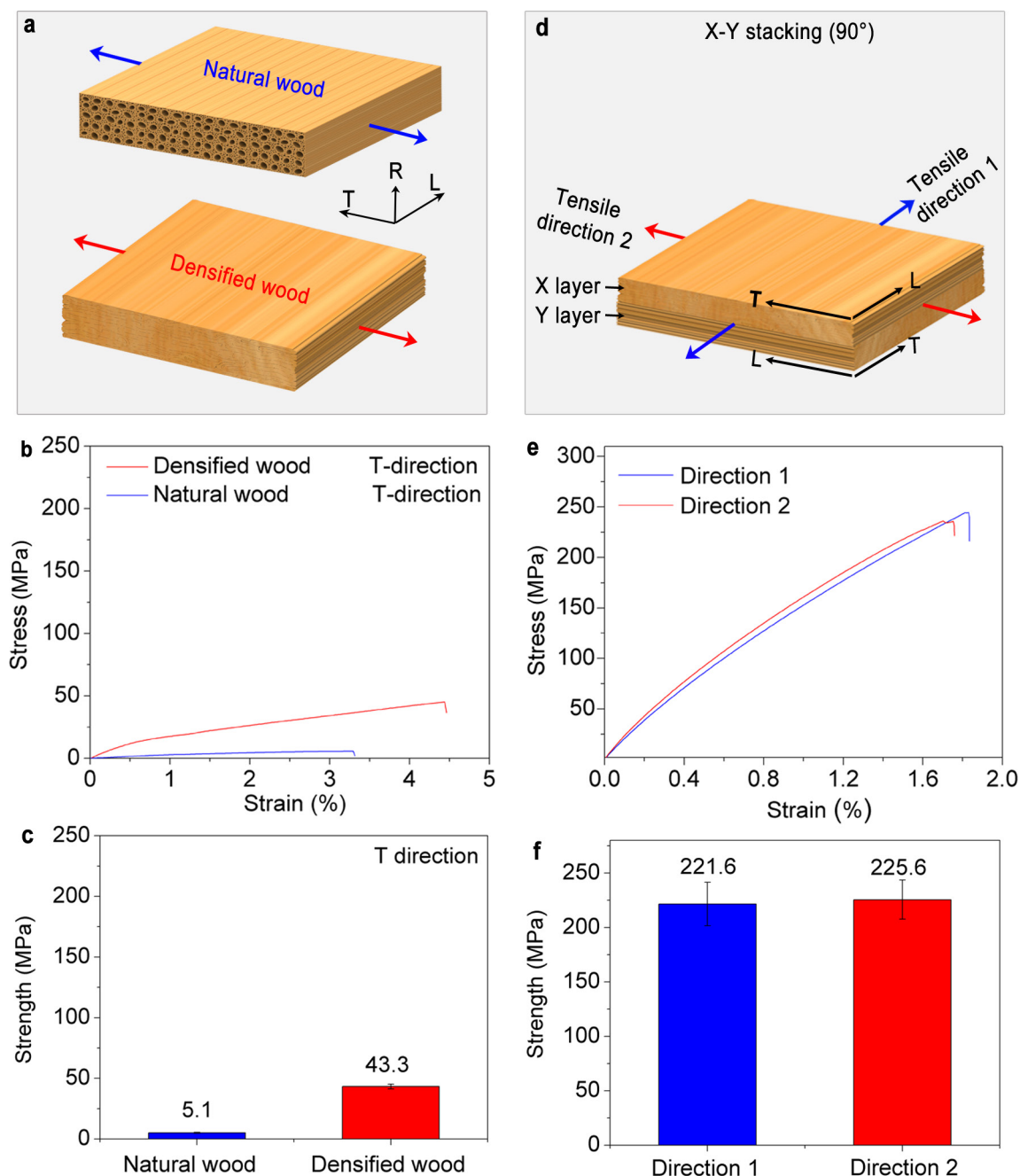
**Extended Data Figure 7 | Fracture surface (RL plane) of natural wood and densified wood.** **a, c,** The schematics of the natural wood and densified wood. **b,** SEM image of the fracture surface of the natural wood showing the pulling out and tearing of the hollow wood lumina along the

fracture surface in the RL plane. **d,** SEM image of the fracture surface of the densified wood in RL plane showing the pulling out and fracture of wood fibres from the densely packed cell walls.



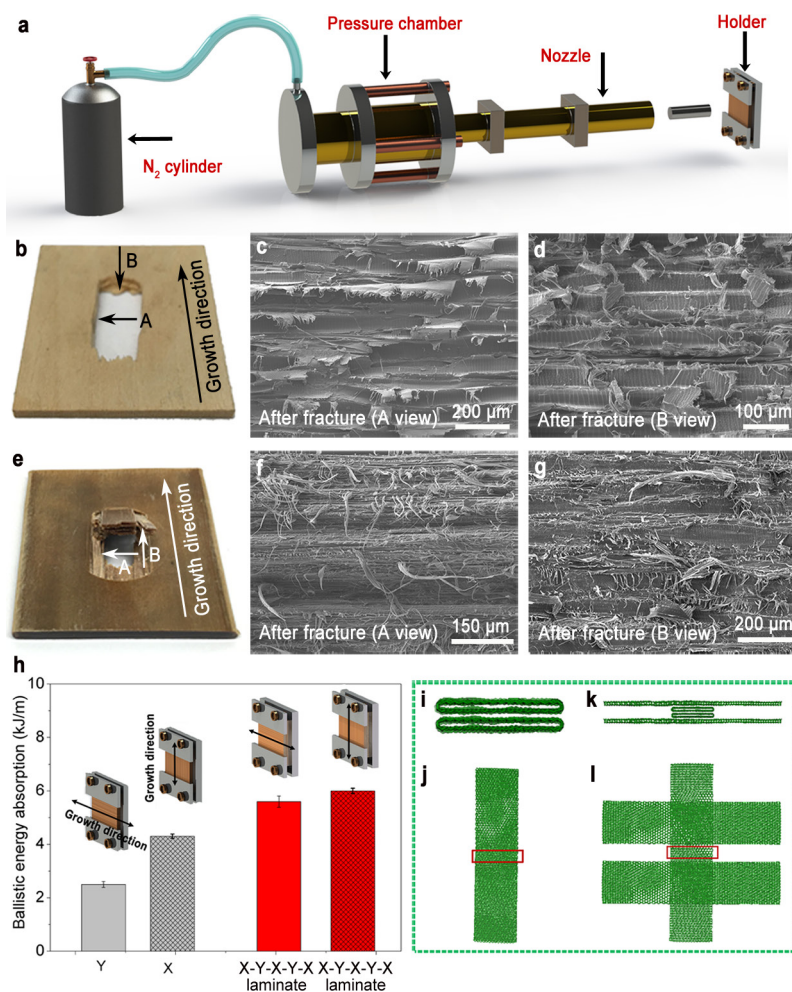
**Extended Data Figure 8 | Simulation model for natural wood and densified wood.** **a**, To obtain the compressed morphology of natural hollow wood fibres in simulations, we apply the reflective wall boundary condition and then gradually shrink one dimension of the simulation box so that the bundle is compressed laterally. **b**, **c**, Morphological view of uncollapsed (**b**) and collapsed (**c**) wood-fibre bundles during pulling. **d**–**f**, Effect of hydrogen bonding (HB). **d**, Simulation model to demonstrate the effect of hydrogen bonding. Two wood fibres slide along each other.

**e**, The corresponding resistant forces with hydrogen bonding turned on and turned off (that is, voiding the Morse potential in the simulation force field) are calculated as a function of sliding displacement, respectively, showing that the hydrogen bonding would increase the resistant force by about ten times. **f**, The initial configuration of the seven-lumina bundle model used in the main text. These lumina each have a diameter of 6.26 nm and length of 8.95 nm. **g**, The values of the coarse-grained parameters used in the simulations.



**Extended Data Figure 9 | Comparison of tensile properties of X-Y stacking densified wood and monolayer densified wood.** **a–c**, Tensile properties of the natural wood and monolayer densified wood along the T direction: **a**, illustration of tensile direction, **b**, tensile stress–strain curves and **c**, tensile strengths along the T direction (natural wood,  $5.1 \pm 0.4$  MPa; densified wood,  $43.3 \pm 2.0$  MPa). **d–f**, Tensile properties of the X–Y stacking densified wood: **d**, illustration of the X–Y stacking

densified wood and two perpendicular tensile directions, **e**, tensile stress–strain curves and **f**, the tensile strengths of the X–Y stacking densified wood along directions 1 and 2 are nearly the same ( $221.6 \pm 20.0$  MPa and  $225.6 \pm 18.0$  MPa, respectively), much higher than that of natural wood and that of monolayer densified wood in the T direction.



**Extended Data Figure 10 | Ballistic test.** **a**, Schematics of the air-gun ballistic tester. **b**, Photograph of natural wood after ballistic test, showing relatively smooth wood surface after the projectile perforates the wood. **c, d**, SEM images of the fracture surface show that fracture takes place along the loosely bonded cell walls in natural wood. **e**, Photograph of monolayer densified wood after ballistic test, showing severely chapped wood surface after the projectile perforates the wood. **f, g**, SEM images of the fracture surface show enormous numbers of wood fibres pulled out from the densely packed cell walls, suggesting substantial energy dissipation during the projectile perforating the densified wood.

**h**, Ballistic energy absorption of the monolayer densified wood (Y,  $2.5 \pm 0.1 \text{ kJ m}^{-1}$ ; X,  $4.3 \pm 0.08 \text{ kJ m}^{-1}$ ) and laminated densified wood (X-Y-X-Y-X laminate:  $5.6 \pm 0.2 \text{ kJ m}^{-1}$ ; X-Y-X-Y-X laminate:  $6.0 \pm 0.1 \text{ kJ m}^{-1}$ ) from both directions (X, fibre alignment direction; Y, perpendicular to fibre alignment direction). The insets show the schematics of the sample and holder. **i-l**, Simulation model used in Fig. 4c, d. **i, j**, End view and top view of the parallel wood fibre model, respectively. **k, l**, End view and top view of the sandwiched wood fibre model, respectively. These wood fibres (before being collapsed) have a diameter of 2.35 nm and a length of 15.34 nm.

# Limited emission reductions from fuel subsidy removal except in energy-exporting regions

Jessica Jewell<sup>1,2</sup>, David McCollum<sup>1,3</sup>, Johannes Emmerling<sup>4,5</sup>, Christoph Bertram<sup>6</sup>, David E. H. J. Gernaat<sup>7,8</sup>, Volker Krey<sup>1</sup>, Leonidas Paroussos<sup>9</sup>, Loïc Berger<sup>4,5,10</sup>, Kostas Fragkiadakis<sup>9</sup>, Ilkka Keppo<sup>11</sup>, Nawfal Saadi<sup>11</sup>, Massimo Tavoni<sup>4,5,12</sup>, Detlef van Vuuren<sup>7,8</sup>, Vadim Vinichenko<sup>13</sup> & Keywan Riahi<sup>1,14</sup>

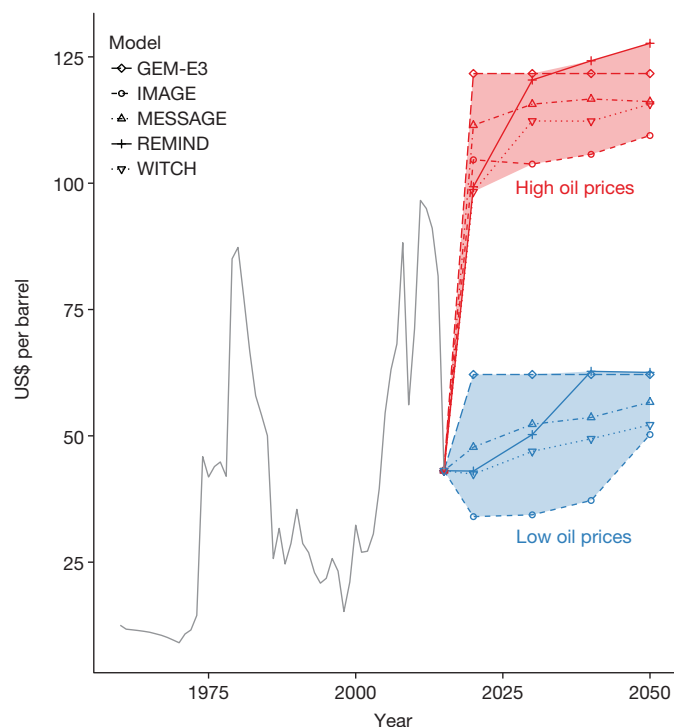
Hopes are high that removing fossil fuel subsidies could help to mitigate climate change by discouraging inefficient energy consumption and levelling the playing field for renewable energy<sup>1–3</sup>. In September 2016, the G20 countries re-affirmed their 2009 commitment (at the G20 Leaders' Summit) to phase out fossil fuel subsidies<sup>4,5</sup> and many national governments are using today's low oil prices as an opportunity to do so<sup>6–9</sup>. In practical terms, this means abandoning policies that decrease the price of fossil fuels and electricity generated from fossil fuels to below normal market prices<sup>10,11</sup>. However, whether the removal of subsidies, even if implemented worldwide, would have a large impact on climate change mitigation has not been systematically explored. Here we show that removing fossil fuel subsidies would have an unexpectedly small impact on global energy demand and carbon dioxide emissions and would not increase renewable energy use by 2030. Subsidy removal would reduce the carbon price necessary to stabilize greenhouse gas concentration at 550 parts per million by only 2–12 per cent under low oil prices. Removing subsidies in most regions would deliver smaller emission reductions than the Paris Agreement (2015) climate pledges and in some regions global subsidy removal may actually lead to an increase in emissions, owing to either coal replacing subsidized oil and natural gas or natural-gas use shifting from subsidizing, energy-exporting regions to non-subsidizing, importing regions. Our results show that subsidy removal would result in the largest CO<sub>2</sub> emission reductions in high-income oil- and gas-exporting regions, where the reductions would exceed the climate pledges of these regions and where subsidy removal would affect fewer people living below the poverty line than in lower-income regions.

Fossil fuel subsidies amounted to about \$330 billion (referring to the US dollar in 2005, throughout) worldwide in 2015 after having reached about \$570 billion in 2013. This fall in subsidies could be partly a sign of reform or simply a reflection of today's lower oil prices, given that historically subsidies have followed the oil price<sup>11</sup> (Supplementary Fig. 1). It is therefore too early to say whether subsidies will continue to fall, stabilize or increase if oil prices rise again. Earlier work found that global subsidy removal by 2020 would reduce greenhouse gas emissions by 5% (ref. 12) to 6% (ref. 13) by 2035 and by 6% (ref. 12) to 8% (refs 14, 15) by 2050. However, all of these studies were done using a single model and none of them explored variations in the oil price, which greatly affects the size of subsidies.

We used five Integrated Assessment Models (IAMs) to evaluate the global and regional effects of removing fossil fuel subsidies on

emissions, the energy mix and energy demand under both low and high oil prices. In the high-oil-price scenarios, oil prices exceed \$100 per barrel and in the low-oil-price scenarios they drop below \$60 per barrel by 2020 (Fig. 1).

The IAMs we use vary in their modelling approaches and solution mechanisms (Supplementary Table 1, Supplementary Information sections 1, 2), which improves the robustness of the results in the face of structural model uncertainties. They include four technology-detailed energy-economy models and one multi-sectoral computable general equilibrium model. An important difference across models, which affects the modelled effects of subsidy removal, is the responsiveness of energy supply and demand to changes in energy prices (Supplementary Tables 1, 2, Supplementary Information sections 2, 3).

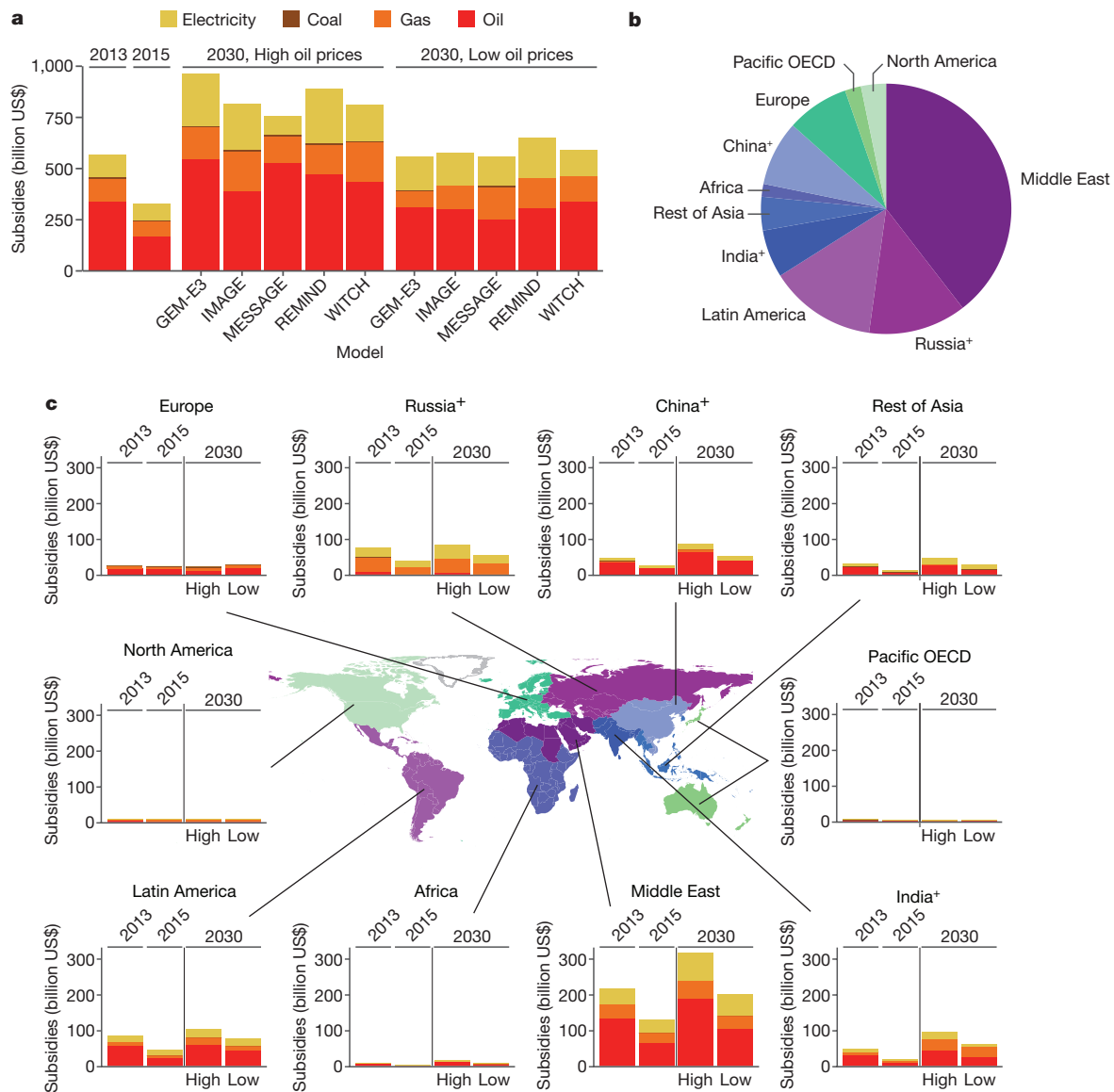


**Figure 1 | Modelled high- and low-oil-price scenarios.** Historical prices represent crude oil prices from ref. 28 and are shown through to the end of 2015. Modelled prices start in 2020.

<sup>1</sup>Energy Program, International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria. <sup>2</sup>Centre for Climate and Energy Transformations and Department of Geography, University of Bergen, 5020 Bergen, Norway. <sup>3</sup>Howard H. Baker Jr. Center for Public Policy, University of Tennessee, Knoxville, Tennessee 37996, USA. <sup>4</sup>Fondazione Eni Enrico Mattei, 20123 Milan, Italy.

<sup>5</sup>Centro Euromediterraneo sui Cambiamenti Climatici, 73100 Lecce, Italy. <sup>6</sup>Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, PO Box 60 12 03, D-14473 Potsdam, Germany. <sup>7</sup>Copernicus Institute for Sustainable Development, University of Utrecht, 3584 CS Utrecht, The Netherlands. <sup>8</sup>PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands. <sup>9</sup>School of Electrical and Computer Engineering, Department of Electric Power, National Technical University of Athens, 15773 Athens, Greece. <sup>10</sup>Department of Economics and Quantitative Methods, IESEG School of Management (LEM-CNRS), 59000 Lille, France. <sup>11</sup>UCL Energy Institute, University College London, London WC1H 0NN, UK. <sup>12</sup>Department of Management, Economics and Industrial Engineering, Politecnico di Milano, 20156 Milan, Italy. <sup>13</sup>Department of Environmental Sciences and Policy, Central European University, 1051 Budapest, Hungary.

<sup>14</sup>Institute of Thermal Engineering, Graz University of Technology, 8010 Graz, Austria.



**Figure 2 | Current and projected fossil fuel subsidies without reform.**

**a**, Global subsidies in 2013 (high oil prices), in 2015 (low oil prices), and in 2030 under high and low oil prices projected in different models. **b**, The regional distribution of subsidies in 2015 (see also Supplementary Table 5). **c**, Subsidies in 2013 and 2015 (Supplementary Table 5) and in

2030 under high and low oil prices in each region (model median).

For model ranges and additional years see Supplementary Tables 5, 7 and 8. The map presents a stylistic representation of regions.

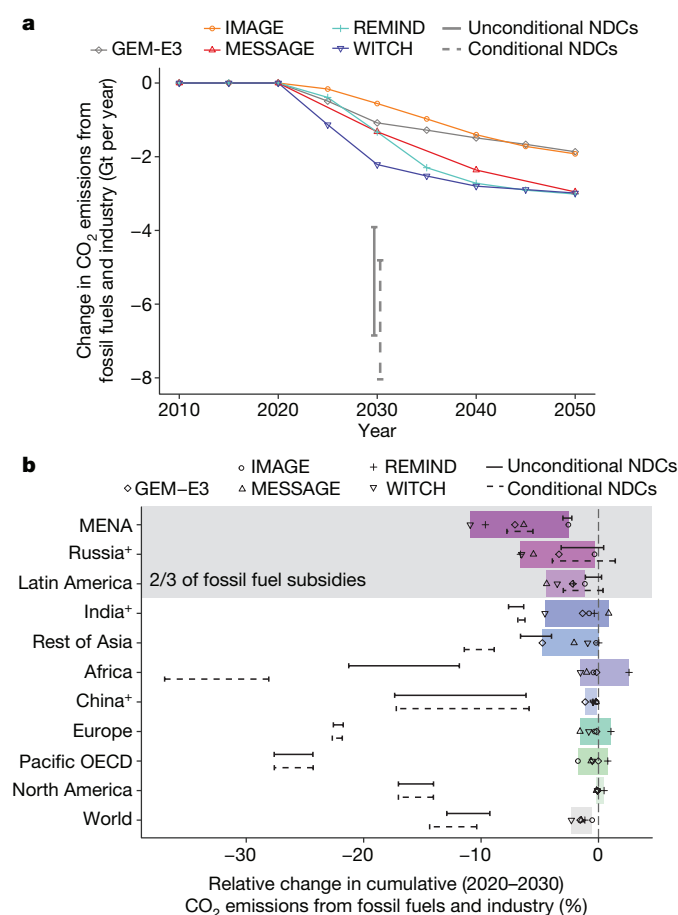
For regional definitions see Supplementary Tables 9–14.

We follow the International Energy Agency (IEA) and the Organisation for Economic Co-operation and Development (OECD) definition of fossil fuel subsidies as government support of the consumption or production of oil, gas or coal that lowers their prices below normal market prices (Methods). This definition excludes un-priced environmental and social externalities such as air pollution and related health effects, which are included in some other estimations<sup>16</sup> but are not appropriate for the purpose of this paper (Methods). We compiled a global comprehensive dataset of fossil fuel subsidies<sup>8,10,11,17,18</sup> under both high and low oil prices (Supplementary Tables 3, 4, Supplementary Information sections 4, 5). In 2013, when oil prices were relatively high, subsidies amounted to approximately \$570 billion (Supplementary Table 5), including \$340 billion for oil, \$110 billion each for natural gas and electricity, and \$5 billion for coal (Fig. 2). Only \$22 billion (less than 4%) were production subsidies (Supplementary Table 3). Following the decline in oil prices, subsidies fell to about \$330 billion in 2015, which amounted to about 10% of energy-related market transactions (Supplementary Table 6).

In our scenarios, we model subsidy rates in a way consistent with historical patterns (Methods). Under high oil prices, by 2030, global subsidies would grow to between \$750 and \$970 billion; under low oil prices, subsidies would be between \$550 and \$700 billion through to the end of 2030 (Supplementary Table 5). In the subsidy removal scenarios, their phase-out starts in 2020 and is completed by 2030.

The three oil- and gas-exporting regions, the Middle East and North Africa (MENA), Russia<sup>+</sup> (the ‘+’ superscript is used to refer to regions that constitute more than only the named country—see Supplementary Table 9 for region definitions) and Latin America accounted for about two-thirds of all fossil fuel subsidies worldwide in 2015 (Fig. 2). In Latin America and MENA, about half of total subsidies goes to oil. In Russia<sup>+</sup>, about half of total subsidies goes to natural gas and the remainder to electricity (mostly generated from natural gas). Of these three regions, subsidy expenditures would grow the most in MENA, which would experience the largest growth in energy use (Fig. 2).

Developing and emerging economies (India<sup>+</sup>, Rest of Asia, Africa and China<sup>+</sup>) currently have lower subsidies than the oil and gas



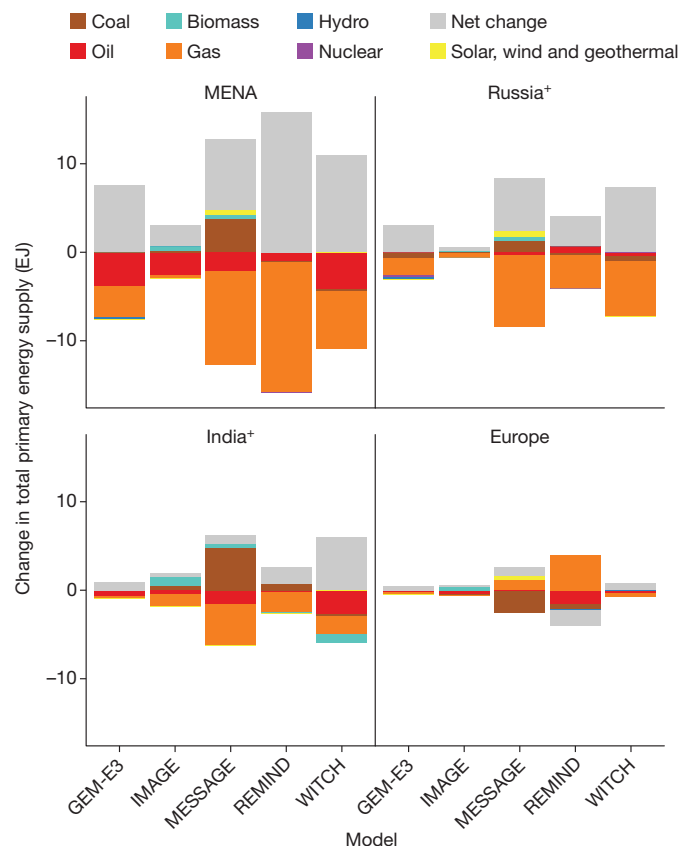
**Figure 3 | Global and regional impact of subsidy removal and NDCs on CO<sub>2</sub> emissions from fossil fuels and industry under low oil prices.**

**a**, The impact of subsidy removal on global annual emissions compared to each model's baseline. **b**, The impact of subsidy removal on cumulative change in emissions from 2020 to 2030 at the regional level (coloured bars). Solid lines represent emission effects of unconditional NDCs and dashed lines of conditional NDCs—both modelled in MESSAGE<sup>29</sup>. The uncertainty ranges for these effects arise from different historical emission inventories, alternative accounting, attribution of non-commercial biomass and uncertainties in the formulations of NDCs (Supplementary Methods, Supplementary Table 15; ref. 29). See Supplementary Fig. 6 for high-oil-price scenarios and Supplementary Fig. 5 for global relative changes and regional absolute changes.

exporters, but their subsidies may grow faster in the future (Fig. 2). Without reform, subsidies in India under high oil prices could become comparable to those in Latin America and Russia<sup>+</sup> by 2030. In these regions, over half of all subsidies goes to oil, for example, through depressed road fuel prices (in countries in the Rest of Asia region), tax breaks on road fuels (in China), or kerosene subsidies (in India and Africa).

Subsidies in the developed regions (Europe, North America and the Pacific OECD) accounted for about 13% of subsidies worldwide in 2015. These are not projected to grow very much in the future.

Subsidy removal would lead to a small decrease in global CO<sub>2</sub> emissions: 0.5–2 gigatons of carbon dioxide (Gt CO<sub>2</sub>) or 1%–4% by 2030 under both low (Fig. 3, Supplementary Fig. 5) and high (Supplementary Figs 5, 6) oil prices. This is much less than the Nationally Determined Contributions (NDCs) from the Paris Agreement, which add up to a decrease of between 4–8 Gt from fossil fuels and industry. Subsidy removal would reduce the average global carbon price in 2020–2050 that would be required to achieve modest climate goals (an atmospheric concentration target of 550 parts per million CO<sub>2</sub> equivalent by 2100 or a probable 2–2.3°C temperature



**Figure 4 | Change in supply of different fuels resulting from subsidy removal in 2030 in four regions under low oil prices.** MENA and Russia<sup>+</sup> illustrate exporting regions, India<sup>+</sup> illustrates developing importing regions, Europe illustrates developed regions (Supplementary Fig. 10 shows the other six regions). Positive values of 'Net change' indicate a decrease in the total primary energy supply; negative values indicate an increase. Supplementary Figs 11 and 12 show results under high oil prices. The regional definitions (Supplementary Tables 9–14) can influence the size of energy system changes.

increase in 2100<sup>19</sup>) by an average of 2%–12% or by \$0.7–\$2.1 per ton of CO<sub>2</sub> under low oil prices (Supplementary Information section 6, Supplementary Tables 16, 17).

Even though the oil price has an impact on the absolute level of subsidies, it does not greatly affect the impact of subsidy removal on emissions because the latter depends on the ratio between subsidies and energy prices, which is similar in the low- and high-oil-price scenarios. Figures 3 and 4 illustrate the low-oil-price scenarios; the high-oil-price scenarios are illustrated in Supplementary Information and described in the text wherever they are very different.

The impacts of subsidy removal are very different in two groups of regions. In oil- and gas-exporting regions (MENA, Russia<sup>+</sup> and Latin America), subsidy removal leads to the largest emission reductions, equivalent to or greater than their relatively modest NDCs. In all other regions, emission reductions from subsidy removal are generally less than their NDCs (Fig. 3, Supplementary Fig. 6).

In Russia<sup>+</sup>, where most subsidies are for natural gas (including electricity generation), subsidy removal would reduce the use of natural gas and generally lead to higher emission reductions than the modest NDCs. In MENA and Latin America, subsidy removal would decrease the use of oil and natural gas leading to emission reductions that are generally comparable to the so-called 'conditional' NDCs (that is, commitments dependent on international action) but generally larger than the unconditional NDCs.

Developing and emerging economies that are not major oil and gas exporters would generally experience smaller emission impacts

(both in absolute terms and in relation to their NDCs) from subsidy removal owing to their lower subsidy levels. The main effect of subsidy removal in India<sup>+</sup> would be reduced use of oil and natural gas, and in the Rest of Asia would be slightly reduced use of coal and oil. In both regions, the decline in emissions would be generally smaller than the NDCs. In China<sup>+</sup> subsidies are lower and the impact of their removal would also be small in comparison with the NDCs. In Africa, subsidy removal would also have a much smaller effect than the NDCs (and in one model would even lead to an increase in emissions owing to the substitution of oil for coal).

In the three developed regions (Europe, North America and the Pacific OECD) with low subsidies, the main impact of global subsidy removal is driven by the change of the price of fossil fuels on the global market. As oil and gas exporters reduce domestic demand by removing subsidies, they make more resources available for the global market. This can, for example, lead to increased use of natural gas in Europe (Fig. 4). This effect is more pronounced in models with more flexible energy trade. The resulting change in emissions can either be negative or positive depending on whether the cheaper natural gas substitutes oil and coal or leads to an increase of consumption. All in all, subsidy removal would lead to much smaller emission reductions than the NDCs.

Although the above results are robust for all models, there are certain variations, due to different features and assumptions of particular models. The most notable difference is that in some regions, subsidy removal can unexpectedly lead to an increase in emissions. In India<sup>+</sup> (the MESSAGE model) and Africa (the REMIND model) this occurs because these models assume more flexibility in fuel substitution. As a result, removing subsidies leads to substitution of oil or natural gas with more carbon-intensive coal, producing either an increase in emissions or smaller reductions of emissions. In addition, REMIND assumes the most flexible international energy trade, which means that energy-importing regions (Europe, the Pacific OECD and North America) increase use of natural gas (and therefore greenhouse gas emissions; Fig. 3) after it stops being subsidized in energy-exporting regions. Other less notable differences are discussed in Supplementary Information section 2.

Our results show that removing fossil fuel subsidies would lower global energy demand. The decrease in energy demand is caused by increasing energy prices and ranges between 5 EJ and 26 EJ per year or 1%–4% in 2030 (Supplementary Figs 7, 8). Under high oil prices, the decrease in demand is larger, reaching up to 30 EJ per year or 7% in 2030 (Supplementary Figs 7, 8). The decrease in demand is largest in oil- and gas-exporting regions (MENA, Russia<sup>+</sup> and Latin America), whereas in some energy-importing regions energy use could even increase following subsidy removal owing to the larger availability of natural gas on international markets (as discussed above).

In addition, removing fossil fuel subsidies would not strongly stimulate the growth of renewable energy by 2030 (Fig. 4). In general, removing fossil fuel subsidies leads to an increase in the share of renewables in regional energy mixes of less than two percentage points (Supplementary Fig. 13). A slightly larger increase may occur under high oil prices in bioenergy in Russia<sup>+</sup>, MENA and Latin America or solar energy in MENA and Russia<sup>+</sup> (Supplementary Figs 10–12). Beyond 2030, subsidy removal could stimulate more noticeable growth of renewable energy, in particular bioenergy under certain modelling assumptions.

A more pronounced effect of fossil subsidy removal is the switch from one fossil fuel to another, for example from subsidized natural gas and oil to coal in MENA, Russia<sup>+</sup> and India<sup>+</sup> as well as from coal and oil to natural gas in Europe (Fig. 4), which highlights the need to consider the systemic effects of subsidy reform policies. The switch between fossil fuels is more pronounced in models with higher flexibility of supply and lower flexibility of demand as well as higher flexibility of international trade (Supplementary Information section 2). Another, more granular effect is the slowdown of the switch from solid

fuels (such as coal and firewood) to natural gas and kerosene among the poor, as shown by IMAGE (a model representing different income groups; see Supplementary Fig. 9). This is in line with earlier findings that as modern fuels become more expensive, lower-income groups are unable to avoid traditional fuels, unless supportive policies are implemented in parallel<sup>20,21</sup>.

We tested the sensitivity of our findings against baseline assumptions (Supplementary Information section 7, Supplementary Figs 14–17), decoupling of the oil and gas prices (Supplementary Information section 8, Supplementary Figs 18–21), and the assumption of higher production subsidies<sup>22,23</sup> (Supplementary Information section 9, Supplementary Table 18, Supplementary Figs 22–25). The emissions and energy systems impacts are generally robust across these uncertainties but changing socio-economic baseline assumptions changes the projected emission reductions from some regional NDCs, which in turn changes the relationship between the NDCs and the effects of subsidy removal (Supplementary Information section 7).

Our finding that subsidy removal would have the largest impact on CO<sub>2</sub> emissions in Russia<sup>+</sup>, MENA and Latin America is especially meaningful when we consider two features of the political economy of subsidies. The first is that subsidy removal could disproportionately harm the poor in some countries<sup>24,25</sup>. The second is that today's low oil prices pressure energy-exporting states to reduce spending as government revenues shrink<sup>24</sup>. This provides a unique political opportunity to remove subsidies precisely where it would have the largest effect on emissions and affect a comparatively small number of people living below \$3.10 per day (Supplementary Table 19, Supplementary Information section 10). Conversely, in low-income regions, subsidy removal would lead to smaller emission reductions and probably affect more people living below the poverty line. The frequently voiced suggestion of coupling subsidy removal with other emission-reduction policies such as carbon pricing<sup>12,15</sup> or clean energy support schemes<sup>26,27</sup> would not necessarily reduce the impact of subsidy removal on the poor unless such policies are specifically designed to do so.

**Data Availability** All data for the subsidy scenarios and sensitivities are available at <https://tntcat.iiasa.ac.at/ADVANCEWP3DB>. The NDC data used in this paper are from ref. 29 and are available on request. The sources and compilation method for the input data on subsidies and prices are described in detail in Supplementary Methods subsection 'Energy price and subsidy data'.

Received 10 January; accepted 11 December 2017.

1. Edenhofer, O. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 32–108 (Cambridge Univ. Press, 2014).
2. International Energy Agency (IEA). *Energy and Climate Change. WEO Special Report 2015*. <https://www.iea.org/publications/freepublications/publication/WEO2015SpecialReportonEnergyandClimateChange.pdf> (OECD/IEA, 2015).
3. Friends of Fossil Fuel Subsidy Reform (FFSR). *Fossil Fuel Subsidy Reform and the Communiqué: Briefing Note*. [http://ffsr.org/wp-content/uploads/2015/06/ffsr\\_information\\_for\\_policymakers-1.pdf](http://ffsr.org/wp-content/uploads/2015/06/ffsr_information_for_policymakers-1.pdf) (FFSR, June 2015).
4. IEA, OPEC, OECD and World Bank. *Joint Report by IEA, OPEC, OECD and World Bank on Fossil-fuel and Other Energy Subsidies: An update of the G20 Pittsburgh and Toronto Commitments*. <https://www.oecd.org/env/49090716.pdf> (2011).
5. G20 Leaders' Communiqué Hangzhou Summit. [https://www.g20.org/profiles/g20/modules/custom/g20\\_beverly/img/timeline/China/2016-09-04-g20-comunique-en.pdf](https://www.g20.org/profiles/g20/modules/custom/g20_beverly/img/timeline/China/2016-09-04-g20-comunique-en.pdf) (2016).
6. International Energy Agency (IEA). *World Energy Outlook 2015*. <http://www.iea.org/publications/freepublications/publication/WEO2015.pdf> (OECD/IEA, 2015).
7. Russell, C. *Oil Price Plunge should Prompt New Fuel Taxes in Asia*: Russell. <https://www.reuters.com/article/us-column-russell-tax-fuel/oil-price-plunge-should-prompt-new-fuel-taxes-in-asia-russell-idUSKBNOKW1V620150123> (Reuters, 2015).
8. International Energy Agency (IEA). *World Energy Outlook 2016*. [https://www.iea.org/bookshop/720-World\\_Energy\\_Outlook\\_2016](https://www.iea.org/bookshop/720-World_Energy_Outlook_2016) (OECD/IEA, 2016).
9. Indonesia's Economy: A good scrap. *Economist* <https://www.economist.com/news/asia/21638179-jokowi-abandons-wasteful-fuel-subsidies-fiscal-prospects-brighten-good-scrap> (8 January 2015).
10. The Organisation for Economic Co-operation and Development (OECD). *OECD Inventory of Support Measures for Fossil Fuels 2015*. <http://www.oecd.org/about/secretary-general/oecd-inventory-of-support-measures-for-fossil-fuels-2015.htm> (OECD, 21 September 2015).

11. International Energy Agency (IEA). *World Energy Outlook 2014*. <https://www.iea.org/publications/freepublications/publication/WE02014.pdf> (OECD/IEA, 2014).
12. Schwanitz, J., Piontek, F., Bertram, C. & Luderer, G. Long-term climate policy implications of phasing out fossil fuel subsidies. *Energy Policy* **67**, 882–894 (2014).
13. International Energy Agency (IEA). *World Energy Outlook 2011*. [https://www.iea.org/publications/freepublications/publication/WE02011\\_WEB.pdf](https://www.iea.org/publications/freepublications/publication/WE02011_WEB.pdf) (OECD/IEA, 2011).
14. Burniaux, J.-M. & Chateau, J. *Mitigation Potential of Removing Fossil Fuel Subsidies*. Paper 853, [http://www.oecd-ilibrary.org/economics/mitigation-potential-of-removing-fossil-fuel-subsidies\\_5kgdx1jr2plp-en](http://www.oecd-ilibrary.org/economics/mitigation-potential-of-removing-fossil-fuel-subsidies_5kgdx1jr2plp-en) (OECD, 2011).
15. Burniaux, J.-M. & Chateau, J. Greenhouse gases mitigation potential and economic efficiency of phasing-out fossil fuel subsidies. *Inter Econ.* **140**, 71–88 (2014).
16. Coady, D., Parry, I., Sears, L. & Shang, B. *How Large are Global Energy Subsidies?* <https://www.imf.org/external/pubs/ft/wp/2015/wp15105.pdf> (International Monetary Fund, 2015).
17. International Energy Agency (IEA). *Energy Subsidies*. <http://www.iea.org/statistics/resources/energysubsidies/> (IEA, accessed 2 May 2017).
18. Wagner, A. *International Fuel Prices 2012/2013*. <https://www.giz.de/expertise/downloads/giz2014-en-international-fuel-prices-2013.pdf> (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2014).
19. Clarke, L. *et al.* in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. *et al.*) 413–510 (Cambridge Univ. Press, 2014).
20. Lam, N. L. *et al.* Kerosene subsidies for household lighting in India: what are the impacts? *Environ. Res. Lett.* **11**, 044014 (2016).
21. Cameron, C. *et al.* Policy trade-offs between climate mitigation and clean cook-stove access in South Asia. *Nat. Energy* **1**, 15010–15015 (2016).
22. Gerasimchuk, I. *et al.* *Zombie Energy: Climate Benefits of Ending Subsidies to Fossil Fuel Production*. <https://www.iisd.org/sites/default/files/publications/zombie-energy-climate-benefits-ending-subsidies-fossil-fuel-production.pdf> (International Institute for Sustainable Development, 2017).
23. Bast, E., Doukas, A., Pickard, S., van der Burg, L. & Whitley, S. *Empty Promises: G20 Subsidies to Oil, Gas and Coal Production*. [http://priceofoil.org/content/uploads/2015/11/empty\\_promises\\_full\\_report\\_update.pdf](http://priceofoil.org/content/uploads/2015/11/empty_promises_full_report_update.pdf) (Oil Change International and Overseas Development Institute, 2015).
24. Inchauste, G. & Victor, D. G. in *The Political Economy of Energy Subsidy Reform* (eds Inchauste, G. & Victor, D. G.) 1–44, <https://openknowledge.worldbank.org/bitstream/handle/10986/26216/9781464810077.pdf> (World Bank, 2017).
25. Lockwood, M. Fossil fuel subsidy reform, rent management and political fragmentation in developing countries. *New Polit. Econ.* **20**, 475–494 (2015).
26. Merrill, L., Bassi, A. M., Bridle, R. & Christensen, L. T. *Tackling Fossil Fuel Subsidies and Climate Change: Levelling the Energy Playing Field*. <http://www.iisd.org/library/tackling-fossil-fuel-subsidies-and-climate-change> (Norden, 2015).
27. Matsuo, T. & Schmidt, T. S. Hybridizing low-carbon technology deployment policy and fossil fuel subsidy reform: a climate finance perspective. *Environ. Res. Lett.* **12**, 014002–014010 (2017).
28. British Petroleum (BP). *BP Statistical Review of World Energy 2016* <https://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-full-report.pdf> (BP, June 2016).
29. Rogelj, J. *et al.* Understanding the origin of Paris Agreement emission uncertainties. *Nat. Commun.* **8**, 15748 (2017).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The research leading to these results received funding from the European Union's Seventh Programme FP7/2007–2013 under grant agreement number 308329 (ADVANCE). We thank the International Institute for Applied Systems Analysis, Energy Program for hosting the online database with the scenario data as well as P. Kolp, L. Groihofer and D. Garcia-Carbrera for data and database support; the International Energy Agency (in particular A. Bromhead, L. Cozzi, N. Selmet, G. Zazias and T. Shirai) for providing data and support related to their energy subsidy database; G. Luderer for contributing to the study design; and A. Cherp for commenting on the manuscript.

**Author Contributions** J.J., D.McC., V.K. and K.R. designed the experiment (with input from C.B. and M.T.). J.J. compiled the fossil fuel subsidies and energy price data. D.McC. and V.K. provided the MESSAGE model data. J.E. and L.B. provided the WITCH model data. D.E.H.J.G. and D.v.V. provided the IMAGE model data. K.F. and L.P. provided the GEM-E3 model data. C.B. provided the REMIND model data. J.J. made all the figures (with assistance from V.V. and D.E.H.J.G.). J.J. led the analysis of the modelling results and writing of the paper, with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.J. ([jewell@iiasa.ac.at](mailto:jewell@iiasa.ac.at)).

**Reviewer Information** Nature thanks H. McJeon, I. Parry and the other anonymous reviewer(s) for their contribution to the peer review of this work.

# Evolutionary history of the angiosperm flora of China

Li-Min Lu<sup>1\*</sup>, Ling-Feng Mao<sup>2\*</sup>, Tuo Yang<sup>1\*</sup>, Jian-Fei Ye<sup>1,3,4\*</sup>, Bing Liu<sup>1,5\*</sup>, Hong-Lei Li<sup>6,7\*</sup>, Miao Sun<sup>8,9\*</sup>, Joseph T. Miller<sup>10,11</sup>, Sarah Mathews<sup>10</sup>, Hai-Hua Hu<sup>1,3</sup>, Yan-Ting Niu<sup>1,3</sup>, Dan-Xiao Peng<sup>1,3</sup>, You-Hua Chen<sup>12</sup>, Stephen A. Smith<sup>13</sup>, Min Chen<sup>14</sup>, Kun-Li Xiang<sup>1,3</sup>, Chi-Toan Le<sup>1,3</sup>, Viet-Cuong Dang<sup>1,3</sup>, An-Ming Lu<sup>1</sup>, Pamela S. Soltis<sup>9</sup>§, Douglas E. Soltis<sup>8,9</sup>§, Jian-Hua Li<sup>15</sup>§ & Zhi-Duan Chen<sup>1,5</sup>§

**High species diversity may result from recent rapid speciation in a ‘cradle’ and/or the gradual accumulation and preservation of species over time in a ‘museum’<sup>1,2</sup>. China harbours nearly 10% of angiosperm species worldwide and has long been considered as both a museum, owing to the presence of many species with hypothesized ancient origins<sup>3,4</sup>, and a cradle, as many lineages have originated as recent topographic changes and climatic shifts—such as the formation of the Qinghai–Tibetan Plateau and the development of the monsoon—provided new habitats that promoted remarkable radiation<sup>5</sup>. However, no detailed phylogenetic study has addressed when and how the major components of the Chinese angiosperm flora assembled to form the present-day vegetation. Here we investigate the spatio-temporal divergence patterns of the Chinese flora using a dated phylogeny of 92% of the angiosperm genera for the region, a nearly complete species-level tree comprising 26,978 species and detailed spatial distribution data. We found that 66% of the angiosperm genera in China did not originate until early in the Miocene epoch (23 million years ago (Mya)). The flora of eastern China bears a signature of older divergence (mean divergence times of 22.04–25.39 Mya), phylogenetic overdispersion (spatial co-occurrence of distant relatives) and higher phylogenetic diversity. In western China, the flora shows more recent divergence (mean divergence times of 15.29–18.86 Mya), pronounced phylogenetic clustering (co-occurrence of close relatives) and lower phylogenetic diversity. Analyses of species-level phylogenetic diversity using simulated branch lengths yielded results similar to genus-level patterns. Our analyses indicate that eastern China represents a floristic museum, and western China an evolutionary cradle, for herbaceous genera; eastern China has served as both a museum and a cradle for woody genera. These results identify areas of high species richness and phylogenetic diversity, and provide a foundation on which to build conservation efforts in China.**

Species composition within a geographic area is the result of historical processes including speciation, extinction, migration<sup>6</sup> and ongoing ecological interactions. The extent to which each process has contributed to spatial and temporal patterns of biodiversity, as well as community assembly, varies across the landscape. The biodiversity patterns within a region may result from a recent increase in the rate of speciation that has generated a cradle of biodiversity. Alternatively, biodiversity may derive from the presence of numerous surviving

ancient lineages, together forming a museum region. The process of speciation and the maintenance of ancient lineages need not be mutually exclusive, and some regions have features of both cradles and museums.

The evolutionary history of regional floras has typically been addressed using specific taxa as exemplars<sup>7–9</sup> or by examining the entire flora at various taxonomic levels<sup>10–12</sup>. These investigations provide insights into historical factors, including geological history, climatic shifts and evolutionary processes, that might have contributed to modern geospatial patterns of biodiversity<sup>13,14</sup>. Concomitantly, these studies lay the foundation for decision-making in conserving biodiversity. However, few studies have explored the biodiversity patterns of a large region incorporating dated phylogenies and detailed distribution data.

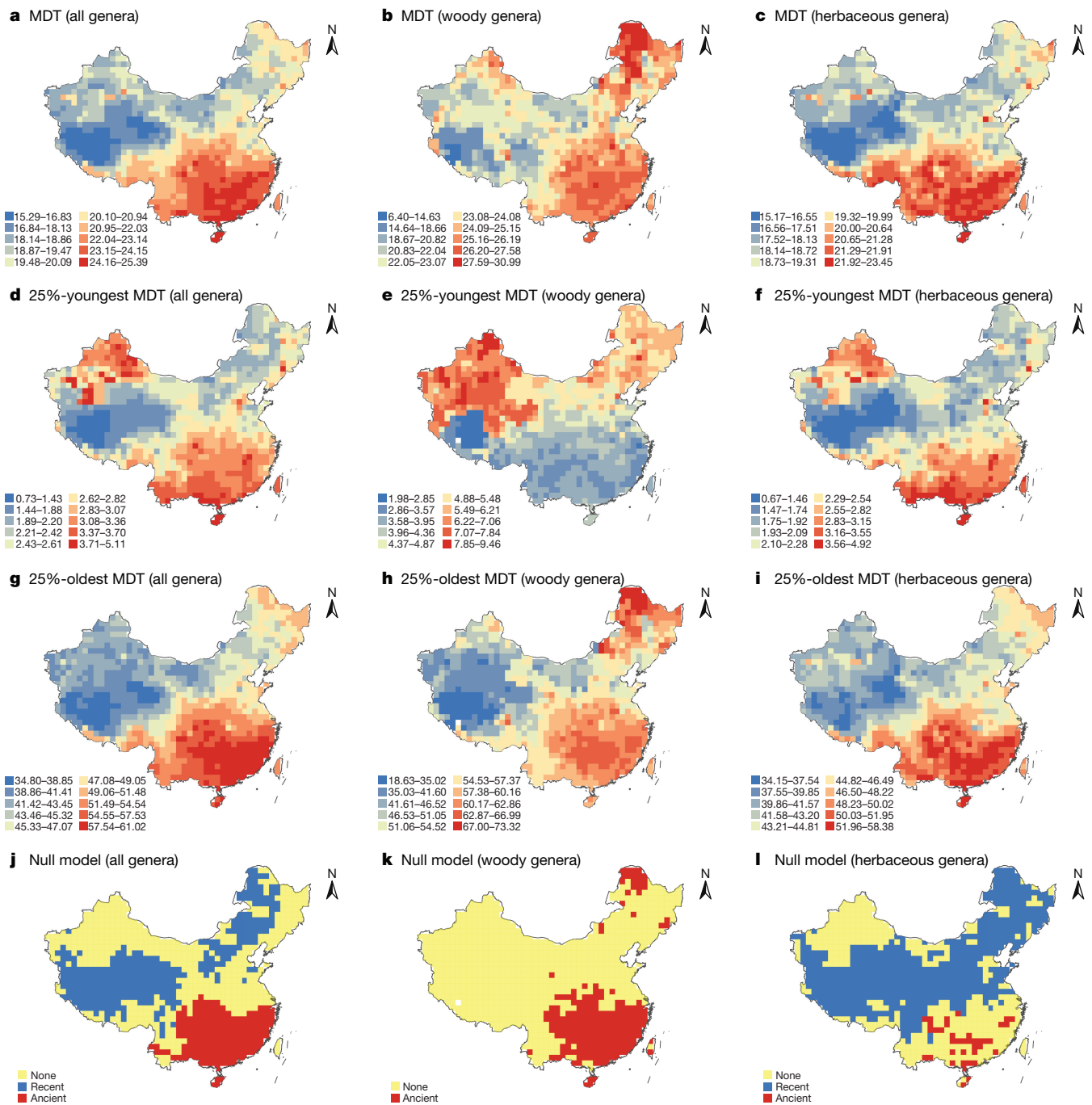
China, which is home to 30,000 of the approximately 350,000–400,000 species of vascular plants<sup>15</sup>, is ideal for investigating patterns of biodiversity because of its large size, range of habitats, considerable biological diversity and heterogeneous physical geography. Whether areas within China serve as cradles or museums remains unclear, as floristic components of putative ancient origin<sup>3,4</sup> and of recent diversification<sup>5</sup> have both been discovered. It has previously been suggested<sup>16</sup> on the basis of comparisons between the taxonomic richness of vascular plants in China and the United States, that the greater species diversity in China reflects the region’s complex topography and long connections with tropical South-East Asia. On the basis of patterns in species richness (using 555 endemic seed plant species), mountainous regions of central and southern China have been identified as the main centres of plant endemism<sup>17</sup>. Previous studies have attributed most of the geographic variation in species richness of woody plants in China to temperature seasonality<sup>18</sup> and the extent of winter cold<sup>19</sup>. Notably, to our knowledge, no previous study has incorporated both phylogenetic and spatial components to address the evolutionary history of the Chinese flora.

We conducted a broad assessment of spatio-temporal divergence patterns and of the assembly of the Chinese angiosperm flora, using a robustly dated phylogeny as well as species distribution data (i) to document the relative proportions of ancient and recent divergences that shaped the extant Chinese angiosperm flora in various geographic regions; (ii) to investigate the differential spatio-temporal divergence patterns of woody and herbaceous genera and their relationships with

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>2</sup>Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China. <sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>4</sup>Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>5</sup>Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China. <sup>6</sup>Chongqing Key Laboratory of Economic Plant Biotechnology/Institute of Special Plants, Chongqing University of Arts and Sciences, Yongchuan 402160, China. <sup>7</sup>Fairylake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China. <sup>8</sup>Department of Biology, University of Florida, Gainesville, Florida 32611-7800, USA. <sup>9</sup>Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA. <sup>10</sup>CSIRO National Research Collections, Australian National Herbarium, Canberra, Australian Capital Territory 2601, Australia. <sup>11</sup>Office of International Science and Engineering, National Science Foundation, Alexandria, Virginia 22314, USA. <sup>12</sup>Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China. <sup>13</sup>Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>14</sup>Institute of Botany, Jiangsu Province & Chinese Academy of Sciences, Nanjing 210014, China. <sup>15</sup>Biology Department, Hope College, Holland, Michigan 49423, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Patterns of the MDTs for Chinese angiosperm genera.** a–i, MDT for all genera, woody genera and herbaceous genera (from left to right), based on all sampled genera (a–c), the youngest 25% of genera (d–f), and the oldest 25% of genera (g–i) in each grid cell. j–l, Null-model test to recognize recent (blue grid cells) and ancient (red grid cells) divergence centres. The analyses included 2,592 angiosperm

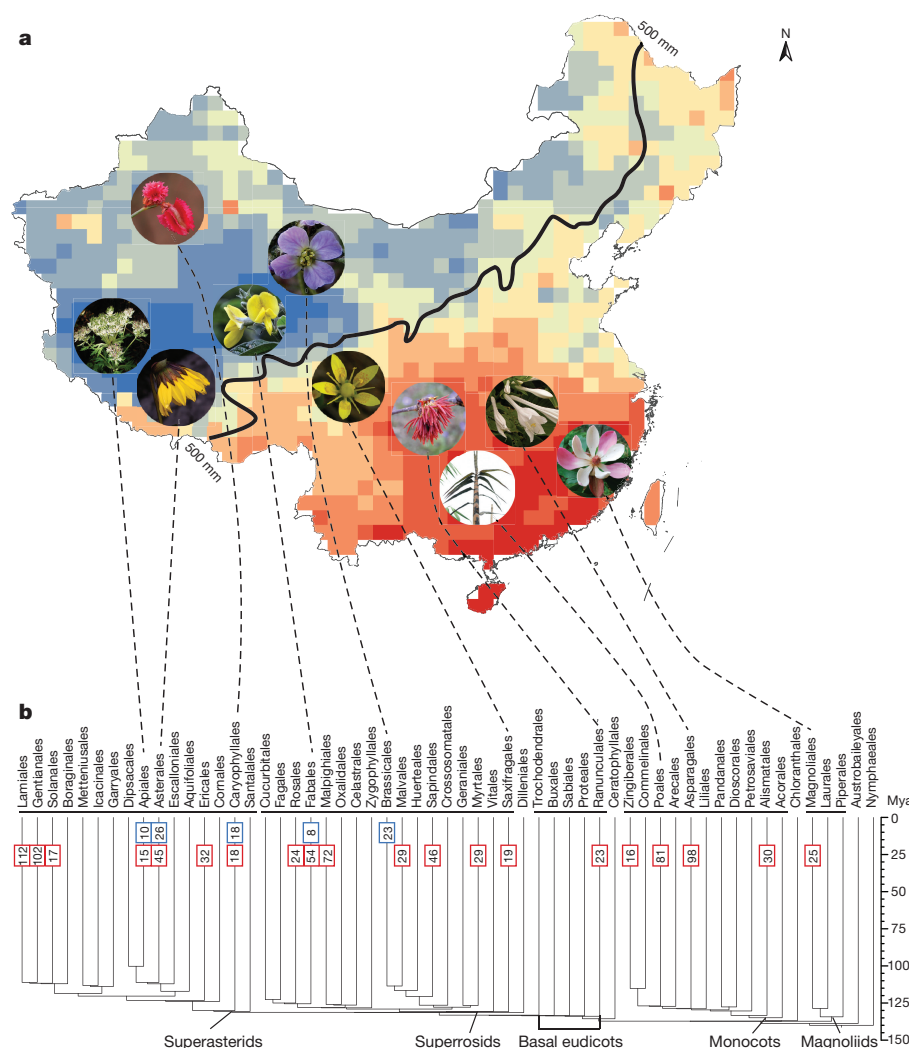
genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

environmental variables; and (iii) to compare genus- and species-level measures of phylogenetic diversity and explore their conservation implications for the Chinese flora.

Our phylogeny resolved evolutionary relationships among all major angiosperm lineages in China (Extended Data Fig. 1), yielding topologies that are highly similar to those for angiosperms as a whole<sup>20,21</sup>. Our estimates of divergence times based on penalized likelihood and PATHd8 are congruent with one another, and agree with those obtained in recent studies of angiosperms on a global basis<sup>22,23</sup> (Extended Data Fig. 2). Divergence time estimates show that 66% of Chinese

angiosperm genera originated during the Neogene and Quaternary periods; the remaining genera diverged in the Palaeogene (29%) and Cretaceous (5%) periods. Additionally, the herbaceous genera have diversified much more rapidly than the woody genera during the past 30 million years (Extended Data Fig. 3).

We divided China into 100-km  $\times$  100-km grid cells, evaluated age variance within grid cells (Extended Data Figs 4, 5), and calculated mean divergence times (MDTs) and median divergence times of genera within each grid cell (Fig. 1; Extended Data Figs 6, 7; Supplementary Information). Mapping the MDTs of all genera revealed a transition belt



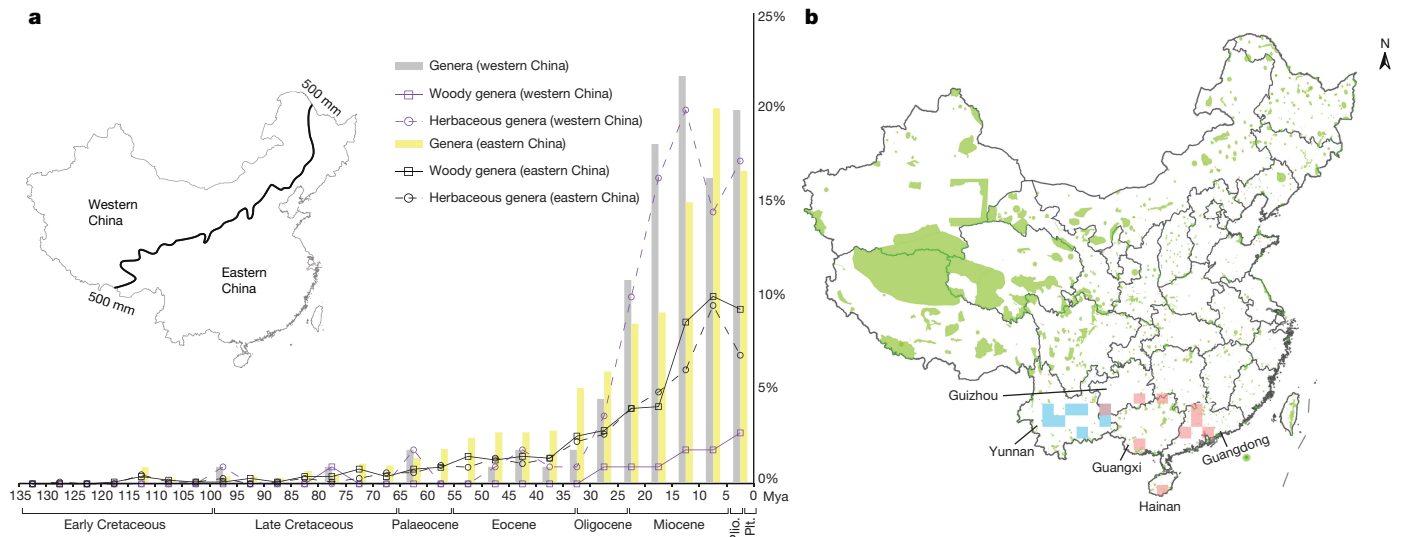
**Figure 2 | Spatio-temporal divergence patterns of the Chinese angiosperm flora.**  
**a**, Patterns of MDTs adapted from Fig. 1a. The dark line represents the 500-mm annual precipitation iseline adapted from ref. 24 (reprinted from ref. 24, with permission from Elsevier). **b**, Ordinal time-tree with the major clades of angiosperms indicated. The top five orders with genera occurring only in western China and top 20 orders with genera occurring only in eastern China are indicated on the tree with blue and red boxes, respectively. The number of genera distributed in western or eastern China from each order is shown within the corresponding box. Map adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

that coincides with the modern 500-mm iseline of annual precipitation, which marks the boundary between humid–semi-humid and arid–semi-arid areas<sup>24</sup> (eastern China versus western China, Fig. 2). Both MDT and null-model analyses indicate that eastern China has older lineages (red grid cells, Fig. 1a, j), particularly in central to southern China. By contrast, western China, and especially the Qinghai–Tibetan Plateau, contains taxa that have diverged more recently (blue grid cells, Fig. 1a, j). Furthermore, our genus-level analyses demonstrate that eastern China is phylogenetically overdispersed with higher phylogenetic diversity, and that western China shows phylogenetic clustering with lower phylogenetic diversity (Extended Data Fig. 8). These findings are also observed in analyses of phylogenetic diversity based on multiple species-level trees, in which taxa that lacked target DNA sequences were provided with meaningful branch lengths using a birth–death clock model (see Methods; Extended Data Fig. 9). The flora of the Cape of South Africa likewise shows phylogenetic structure—the western region is phylogenetically clustered, and the eastern region is overdispersed<sup>10</sup>. However, taxon richness is decoupled from phylogenetic diversity in the Cape of South Africa; in China, taxon richness and phylogenetic diversity are positively correlated.

Western China includes the arid north-western portion of the country and most of the Qinghai–Tibetan Plateau (Fig. 2). A fundamental climate shift may have occurred in western China as recently as the early Miocene, owing to the uplift of the Qinghai–Tibetan Plateau and subsequent development of the Asian monsoon<sup>24,25</sup>. Of the 111 genera that occur only in western China, 76% originated in the past 20 million years and only 24% originated before this time. In western China, a much higher percentage of herbaceous than woody

genera has originated since 30 Mya (Fig. 3a). Moreover, genera that occur only in western China are predominantly members of only a few clades (Apiales, Asterales and Brassicales), most of which have much younger divergence times than the major clades of eastern China (Fig. 2; Extended Data Table 1). MDTs calculated from the youngest 25% of herbaceous genera in each grid cell also indicate that western China—in particular the Qinghai–Tibetan Plateau—has younger lineages (Fig. 1f) than eastern China, which further suggests that western China represents a cradle for herbaceous angiosperms.

Mountainous areas of eastern China have been proposed as refugia for plants that originated in the early Cretaceous or late Jurassic periods<sup>26,27</sup> because their geological environment and climate (including orogenic movements, annual temperature and annual precipitation) may have experienced little change since the Cretaceous<sup>28</sup>. Of the 1,026 genera that occur only in eastern China, 39% originated before 20 Mya and 61% arose more recently than this. Both herbaceous and woody genera diverged at similar rates throughout geological time (Fig. 3a). The 20 major clades with the largest number of genera occurring only in eastern China are distributed throughout the ordinal-level time-tree from early-diverging clades (for example, Alismatales, Asparagales, Magnoliales and Ranunculales) to later-diverging lineages (for example, Asterales, Gentianales and Lamiales) (Fig. 2; Extended Data Table 1). MDTs based on the youngest 25% and oldest 25% of genera in each grid cell reveal that eastern China has old herbaceous lineages (Fig. 1f, i), but has both old and young woody lineages (Fig. 1e, h). Eastern China may have served as a museum for herbaceous genera, but as both a museum and a cradle for woody genera.



**Figure 3 | Angiosperm divergence pattern and conservation priorities in western and eastern China.** **a**, Percentage of genera occurring only in western ( $n = 111$ ) or eastern China ( $n = 1,026$ ) during geological time. Western China has a higher percentage of herbaceous genera (purple dashed line) than woody genera (purple solid line) that have originated since 20 Mya. Western and eastern China are divided by the 500-mm

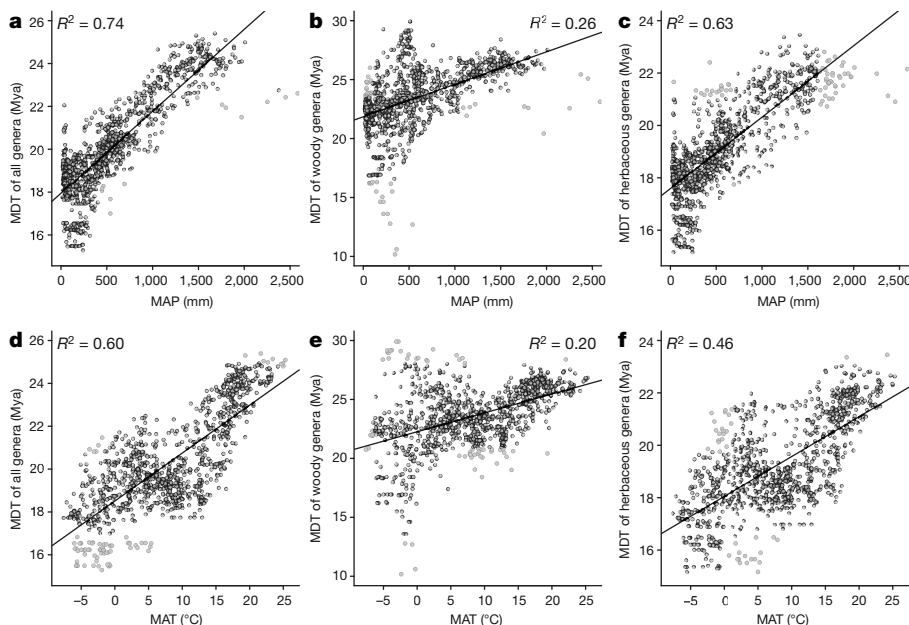
isoline of annual precipitation. Plio., Pliocene epoch; Pli., Pleistocene epoch. **b**, Grid cells with the top 5% highest phylogenetic diversity and SES-PD at genus (pink) and species (blue) levels. Protected areas are highlighted in green. Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

The mean annual precipitation (MAP) and mean annual temperature (MAT) have higher explanatory power for the MDTs of the herbaceous genera (Fig. 4c, f) than of the woody genera (Fig. 4b, e). These patterns may reflect the heterogeneity in rates of evolution between herbaceous and woody lineages. Herbaceous plants are well known to have higher substitution rates owing to their shorter generation times, which perhaps allows them to respond more quickly to environmental change through increased genetic divergence and speciation rates<sup>23,29</sup>.

The spatial divergence and diversity patterns of angiosperms detected here do not precisely reflect the latitudinal gradient in China; MDT and phylogenetic diversity decrease from south-east to north-west (Fig. 2a; Extended Data Fig. 8d, g). Our results show the importance of water and temperature in limiting the dispersal of species from humid and warm regions to drier and colder areas. The effects of topography, with a pronounced altitudinal gradient increasing from east to west, and the monsoon climate in eastern Asia are so extensive that the

decreasing temperature and precipitation gradients from south-eastern to north-western China are not consistent with the latitudinal gradient, as might be expected in flatter regions.

On the basis of a species-level phylogenetic tree and distribution data with 'county' as the basic unit, we inferred that the species richness and phylogenetic diversity in protected areas cover approximately 88% and 96%, respectively, of the total species richness and phylogenetic diversity in China. For conservation planning, these values may be over-estimates that result from the coarse scale of our distributional data, as most nature reserves are smaller in size than Chinese counties. Notably, areas with the top 5% highest phylogenetic diversity and standard effective size of phylogenetic diversity (SES-PD) are mainly located in several provinces of eastern China (Fig. 3b): Guangdong, Guangxi, Guizhou and Hainan for genus-level phylogenetic diversity hotspots, and Yunnan for species-level phylogenetic diversity. These areas are also hotspots for threatened plants in China<sup>30</sup>. However, in contrast



**Figure 4 | Regression analyses between MDT and two environmental variables for the Chinese angiosperm genera.** **a–c**, MDT and MAP. **d–f**, MDT and MAT. From left to right, patterns for all genera (**a**, **d**), woody genera (**b**, **e**) and herbaceous genera (**c**, **f**). The analyses were conducted across all grid cells ( $n = 943$ ) and used the non-spatial linear regression model.

to western China, protected areas in eastern China are fragmented (Fig. 3b), largely as a result of urbanization and administrative division. Our data suggest the need to establish more connections between existing nature reserves and national parks that span provincial borders to conserve plant lineages of ancient and recent origins in eastern China, as well as the other organisms that depend on these floristic elements. These findings should be of broad interest to evolutionary and conservation biologists, and serve to stimulate better-informed conservation planning and research.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 26 January; accepted 22 December 2017.**

**Published online 31 January 2018.**

- McKenna, D. D. & Farrell, B. D. Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. *Proc. Natl Acad. Sci. USA* **103**, 10947–10951 (2006).
- Moreau, C. S. & Bell, C. D. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* **67**, 2240–2257 (2013).
- Blackmore, S., Hong, D.-Y., Raven, P. H. & Wortley, A. H. in *Plants of China: A Companion to the Flora of China* (eds Hong, D.-Y. & Blackmore, S.) 1–6 (Science Press, 2013).
- Sun, G., Dilcher, D. L., Zheng, S.-L. & Zhou, Z.-K. In search of the first flower: a Jurassic angiosperm, *Archaeofructus*, from northeast China. *Science* **282**, 1692–1695 (1998).
- Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong, Y. & Sun, H. Evolutionary diversifications of plants on the Qinghai–Tibetan Plateau. *Front. Genet.* **5**, 4 (2014).
- Xing, Y.-W. & Ree, R. H. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc. Natl Acad. Sci. USA* **114**, E3444–E3451 (2017).
- Klak, C., Reeves, G. & Hedderson, T. Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature* **427**, 63–65 (2004).
- Mishler, B. D. *et al.* Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. *Nat. Commun.* **5**, 4473 (2014).
- Richardson, J. E., Pennington, R. T., Pennington, T. D. & Hollingsworth, P. M. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**, 2242–2245 (2001).
- Forest, F. *et al.* Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760 (2007).
- Verboom, G. A. *et al.* Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? *Mol. Phylogenet. Evol.* **51**, 44–53 (2009).
- Thornhill, A. H. *et al.* Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *J. Biogeogr.* **43**, 2085–2098 (2016).
- Donoghue, M. J. A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl Acad. Sci. USA* **105**, 11549–11555 (2008).
- Svenning, J.-C., Eiserhardt, W. L., Normand, S., Ordóñez, A. & Sandel, B. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **46**, 551–572 (2015).
- Wu, Z.-Y., Raven, P. H. & Hong, D.-Y. (eds) *Flora of China*, Vol. 1–25 (Science Press & Missouri Botanical Garden Press, 1994–2013).
- Qian, H. & Ricklefs, R. E. A comparison of the taxonomic richness of vascular plants in China and the United States. *Am. Nat.* **154**, 160–181 (1999).
- López-Pujol, J., Zhang, F.-M., Sun, H.-Q., Ying, T.-S. & Ge, S. Centres of plant endemism in China: places for survival or for speciation? *J. Biogeogr.* **38**, 1267–1280 (2011).
- Qian, H. Environmental determinants of woody plant diversity at a regional scale in China. *PLoS ONE* **8**, e75832 (2013).
- Wang, Z.-H., Fang, J.-Y., Tang, Z.-Y. & Lin, X. Patterns, determinants and models of woody plant diversity in China. *Proc. R. Soc. Lond. B* **278**, 2122–2132 (2011).
- The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
- Soltis, D. E. *et al.* Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **98**, 704–730 (2011).
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
- Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
- Sun, X.-J. & Wang, P.-X. How old is the Asian monsoon system?—Palaeobotanical records from China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **222**, 181–222 (2005).
- Favre, A. *et al.* The role of the uplift of the Qinghai–Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev. Camb. Philos. Soc.* **90**, 236–253 (2015).
- Axelrod, D. I., Al-Shehbaz, I. A. & Raven, P. H. in *Floristic Characteristics and Diversity of East Asian Plants* (eds Zhang, A.-L. & Wu, S.-G.) 43–55 (China Higher Education, 1996).
- Benton, M. J. *The Fossil Record 2* (Chapman & Hall, 1993).
- Wu, Z.-Y., Sun, H., Zhou, Z.-K., Li, D.-Z. & Peng, H. *Floristics of Seed Plants from China* (Science Press, 2010).
- Smith, S. A. & Beaulieu, J. M. Life history influences rates of climatic niche evolution in flowering plants. *Proc. R. Soc. Lond. B* **276**, 4345–4352 (2009).
- Zhang, Z.-J., He, J.-S., Li, J.-S. & Tang, Z.-Y. Distribution and conservation of threatened plants in China. *Biol. Conserv.* **192**, 454–460 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J.-Y. Fang, D.-Z. Li, K.-P. Ma, S.-Z. Zhang, H. Sun, J.-Q. Liu, Z.-H. Wang, X.-Q. Wang and H.-Z. Kong for help initiating this study. This research was supported by the National Key Basic Research Program of China (2014CB954100), the National Natural Science Foundation of China (31590822), the Chinese Academy of Sciences International Institution Development Program (SAJC201613), the National Natural Science Foundation of China and US National Science Foundation Dimensions Collaboration Project (31461123001), the US National Science Foundation (Open Tree of Life: DEB-1207915, DEB-1208428; ABI DBI-1458466 and DBI-1458640; iDigBio: EF-1115210 and DBI-1547229; US–China Dimensions of Biodiversity: DEB-1442280) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

**Author Contributions** Z.-D.C., P.S.S., D.E.S. and J.-H.L. conceived the paper. L.-M.L., L.-F.M., T.Y., J.-F.Y., B.L., H.-L.L. and M.S. analysed the data. L.-M.L., L.-F.M., T.Y., J.-F.Y., B.L., J.T.M., S.M., P.S.S., D.E.S., J.-H.L. and Z.-D.C. wrote the first draft and finalized the manuscript. H.-H.H., Y.-T.N., D.-X.P., M.C., K.-L.X., C.-T.L. and V.-C.D. contributed data. J.T.M., A.-M.L., Y.-H.C., S.A.S., P.S.S., D.E.S., J.-H.L. and Z.-D.C. contributed substantially to revisions. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to Z.-D.C. ([zhidian@ibcas.ac.cn](mailto:zhidian@ibcas.ac.cn)).

**Reviewer Information** Nature thanks R. Colwell, V. Savolainen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Phylogeny reconstruction.** Sequences of four plastid genes (*atpB*, *matK*, *ndhF* and *rbcL*) and one mitochondrial gene (*matR*) were used to reconstruct the phylogeny of Chinese vascular plants<sup>31</sup>. Generic circumscriptions were based on ref. 15. We sampled one species for the 1,173 genera with only one species in China. For the 1,736 genera with 2–30 species in China, two species were sampled from each genus. For the 267 genera with more than 30 species in China, approximately 10% of the species of each genus were sampled, reflecting intrageneric diversity. We downloaded all available sequences for the target DNA regions from GenBank; if more than one sequence was available for the same locus for a species, the longest one of good quality was selected. For genera with sequences that were unavailable in the public database (781 genera in total), we generated new sequences from leaf materials, collected from the field for 513 genera and from specimens from the Chinese National Herbarium (PE) for 47 genera. There are 231 genera that remain unavailable because we failed to obtain the materials or amplify the target sequences. Details of DNA extraction, PCR, sequencing, alignment, accession numbers of sequences and phylogeny reconstruction have previously been published<sup>31</sup>.

**Divergence time estimation.** We used the penalized likelihood method as implemented in treePL<sup>32</sup> (<https://github.com/blackrim/treePL>) to date divergence times of Chinese angiosperms based on the optimal maximum likelihood phylogram obtained with RAXML 8.0.22<sup>33</sup> in the CIPRES Science Gateway<sup>34</sup>, after excluding the outgroups (for example, lycophytes, ferns and gymnosperms). Our dated phylogeny included 5,864 species native to China, representing 2,665 genera from 273 families or approximately 92% of the angiosperm genera of China. We validated the available fossils and selected 138 calibrations for dating analyses (Supplementary Table 1 in Supplementary Information). The 'prime' option was applied to identify the best optimization parameters, and a 'thorough' analysis was then carried out with the optimal parameters determined above (opt = 1, optad = 1 and optcvad = 4). To identify the best smoothing parameter that affects the penalty for rate variation over the phylogram, a 'random subsample and replicate cross-validation' was conducted with treePL. Confidence intervals for each node were calculated following previously published methods<sup>22</sup>. To accommodate for variation in branch length estimates, we calculated 100 bootstrap replicates with topology fixed to the above maximum likelihood phylogram but with varying branch lengths. We then conducted treePL on these 100 replicates. Age statistics for all nodes were summarized with TreeAnnotator v.1.8.4<sup>35</sup>.

We also used an alternative dating method, PATHd8<sup>36</sup>, to estimate divergence times of Chinese angiosperms. The calibrations for the PATHd8 analysis were identical to those used for the treePL analysis, except that the crown age of angiosperms was set to 138 Mya instead of a maximum of 140 Mya and a minimum of 136 Mya (as in treePL) because PATHd8 requires one fixed calibration. Both treePL and PATHd8 are rate-smoothing methods, but PATHd8 sequentially takes averages over path lengths from an internode to all its descending terminals, one pair of sister groups at a time<sup>37</sup>, where smoothing is done stepwise for each node separately; by contrast, smoothing in treePL is done simultaneously over the tree. The correlation between ages at all nodes based on the treePL and PATHd8 analyses was assessed with Spearman's rank correlation analysis in R v.3.2.0<sup>38</sup>.

To evaluate whether dates for this regional time-tree are biased owing to the geographic sampling, we used a correlation analysis to compare our estimated divergence times with recent global-scale angiosperm time-tree reconstructions<sup>22,23</sup>; one of these represents a family-level time-tree with multiple fossil calibration points<sup>22</sup>, and the other is a species-level time-tree with dense taxon sampling (32,223 species) and fewer calibrations<sup>23</sup>. The stem age of each family was extracted for the Spearman's rank correlation analyses. Only the family ages were compared (circumscription of families, following ref. 20), because different genera and species were included in the three studies. Ages of genera were extracted from our dichotomous time-tree estimated by treePL for the downstream analyses. For monophyletic genera, stem ages were extracted directly by tracing their stem node. For genera that are polyphyletic or paraphyletic (380 out of 2,665), the stem age of each monophyletic lineage was extracted and the oldest one was selected as the age of the genus. The numbers of angiosperm genera that originated during specified geological timespans are provided in Extended Data Fig. 3, with the global temperature changes since 65 Mya<sup>39</sup> indicated.

**Distribution of angiosperm species in China.** The spatial distribution data and information on growth form were assembled from nearly all published national and provincial floras, as well as some local floras, checklists and herbarium records. The spatial distribution data are at the county level (2,377 counties) with an average county-size of approximately 4,000 km<sup>2</sup>. To minimize the sampling bias of unequal sampling areas, we divided the map of China into 100-km × 100-km grid cells, and grid cells on the border that cover less than 50% of the area of a grid cell (that is, 5,000 km<sup>2</sup>) were excluded from the analyses. Maps of China used in this study were

adapted from standard maps released by the National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576). The gridded distribution database contained 1,409,239 occurrence records for 26,978 angiosperm species from 2,845 genera. After matching with the phylogeny, the final dataset included a total of 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ).

#### Spatial distribution of MDTs and null-model test for divergence hotspots.

To explore the spatial divergence patterns of Chinese angiosperm genera, we calculated the weighted MDTs of all genera in each grid cell by integrating spatial distribution data with our dated phylogenetic tree. AGE<sub>*i*</sub> represented the age of a genus *i* ( $i = 1, \dots, n$ ) in a grid cell, and *S<sub>i</sub>* the species number in genus *i* in this grid cell. From this, MDT was calculated as:

$$\text{MDT} = \frac{(\text{AGE}_1 \times S_1) + (\text{AGE}_2 \times S_2) + (\text{AGE}_3 \times S_3) + \dots + (\text{AGE}_n \times S_n)}{S_1 + S_2 + S_3 + \dots + S_n}$$

We further divided the genus dates in each grid cell into quartiles and calculated MDTs on the basis of the youngest and oldest quartiles, separately, in each grid cell. The MDTs based on the youngest quartile allowed us to recognize centres of recent divergence, whereas MDTs based on the oldest quartile detected ancient centres of divergence. To avoid potential bias from grid cells that had either relatively old or young genera, we ranked all genera from youngest to oldest, partitioned them into quartiles based on their ages, computed MDT in each cell for the absolute youngest 25% and the absolute oldest 25% of genera, and then mapped the results across China.

We designed a null model to identify ancient and recent divergence hotspots for the angiosperm flora of China. The mean ages of the youngest and oldest quartiles in each grid cell were selected as the observed values for the null models, and then we shifted the genera randomly using all genera investigated in China as a sampling pool to obtain the null distributions of ages for the youngest and oldest quartiles for each grid cell. The standardized effect size of the mean divergence time (SES-MDT) of genera for each grid cell was calculated as:

$$\text{SES-MDT} = \frac{\text{MDT}_{\text{observed}} - \text{MDT}_{\text{random}}}{\text{s.d.}(\text{MDT}_{\text{random}})}$$

where MDT<sub>observed</sub> is the observed MDT; MDT<sub>random</sub> is the expected MDT of the randomized assemblages ( $n = 999$ ); and s.d.(MDT<sub>random</sub>) is the s.d. of the MDT for the randomized assemblages. Grid cells with values of SES-MDT for the youngest quartile below −1.96 were identified as notable hotspots of recent divergence, whereas grid cells with SES-MDT for the oldest quartile above 1.96 were identified as notable hotspots of ancient divergence. Considering that the evolutionary history of herbaceous and woody plants may differ<sup>40</sup>, the above analyses were also conducted separately for herbaceous and woody genera. Analyses of MDT were implemented in R and ArcGIS 10.1 (<http://www.esri.com/>).

Previous studies have demonstrated that the overall species richness patterns of birds are largely determined by the geographically wide-ranging species<sup>41–43</sup>, indicating that patterns may be driven by a subset of taxa and may not be representative of an entire biota. To explore whether MDT patterns for China are influenced largely by values for widespread species, we ranked genera from the most restricted to most widespread in China, partitioned the genera into quartiles on the basis of their range size and mapped MDT for each quartile following a previously published description<sup>41</sup>.

**Spatial distribution of median divergence times.** Age variation within grid cells was evaluated by plotting divergence times in each grid cell (Extended Data Fig. 4) and calculating the skewness and kurtosis of divergence times (Extended Data Fig. 5). To verify the results of MDT, we also investigated the distribution patterns of the Chinese angiosperm genera by mapping the median divergence times (medianDT) based on all genera, and the youngest and oldest quartiles in each grid cell. The null model for the median divergence time applied a modified effective-size statistic<sup>44–46</sup> and was calculated as:

$$\text{SES-medianDT} = \frac{\text{medianDT}_{\text{observed}} - \text{medianDT}_{\text{random}}}{1.4826 \times \text{MAD}_{\text{random}}}, \text{ if } \text{MAD}_{\text{random}} > 0$$

$$\text{SES-medianDT} = \frac{\text{medianDT}_{\text{observed}} - \text{medianDT}_{\text{random}}}{1.2553 \times \text{meanAD}_{\text{random}}}, \text{ if } \text{MAD}_{\text{random}} = 0$$

where medianDT<sub>observed</sub> is the observed median divergence time; medianDT<sub>random</sub> is the expected median divergence time of the randomized assemblages ( $n = 999$ ); MAD<sub>random</sub> is the median absolute deviation of the divergence times for the randomized assemblages; and meanAD<sub>random</sub> is the mean absolute deviation of the divergence times for the randomized assemblages.

**Richness and phylogenetic diversity.** We calculated the generic richness, Faith's phylogenetic diversity<sup>47</sup> and SES-PD of the Chinese angiosperm genera on the basis of our ultrametric chronogram using the 'picante' package in R. Faith's phylogenetic diversity is the sum of all phylogenetic branch lengths that connect species in a community. We calculated phylogenetic diversity as the length of the subtree that joins the genera in each grid cell to the root of the chronogram. SES-PD was calculated because phylogenetic diversity is usually positively correlated with species richness<sup>48</sup>. We first obtained a null distribution of the expected phylogenetic diversity values by shuffling taxa labels across the tips of the tree 999 times for each grid cell. SES-PD was then calculated by dividing the difference between the observed (PD<sub>observed</sub>) and expected phylogenetic diversity (PD<sub>random</sub>) by the s.d. of the null distribution (s.d.(PD<sub>random</sub>)):

$$\text{SES-PD} = \frac{\text{PD}_{\text{observed}} - \text{PD}_{\text{random}}}{\text{s.d.}(\text{PD}_{\text{random}})}$$

**Phylogenetic structure.** The net relatedness index (NRI) and the nearest taxon index (NTI) were calculated to investigate the phylogenetic structure (clustering or overdispersion) of angiosperm genera across China<sup>49</sup>. NRI is based on the mean phylogenetic distance (MPD), an estimate of the average phylogenetic relatedness between all possible pairs of taxa within a grid cell, and primarily reflects structure at deeper parts of the phylogeny. NTI is based on mean nearest taxon distance (MNTD), an estimate of the mean phylogenetic relatedness between each pair of taxa in a grid cell and its nearest relative in the phylogeny, and reflects shallower parts of the phylogeny. NRI and NTI were calculated as follows:

$$\text{NRI} = -1 \times \frac{\text{MPD}_{\text{observed}} - \text{MPD}_{\text{random}}}{\text{s.d.}(\text{MPD}_{\text{random}})}$$

$$\text{NTI} = -1 \times \frac{\text{MNTD}_{\text{observed}} - \text{MNTD}_{\text{random}}}{\text{s.d.}(\text{MNTD}_{\text{random}})}$$

where MPD<sub>observed</sub> and MNTD<sub>observed</sub> are the observed MPD and MNTD; MPD<sub>random</sub> and MNTD<sub>random</sub> are the averages of the expected MPD and MNTD of the randomized assemblages ( $n = 999$ ); and s.d.(MPD<sub>random</sub>) and s.d.(MNTD<sub>random</sub>) are the standard deviation of MPD<sub>random</sub> and MNTD<sub>random</sub> for the randomized assemblages. The null distributions of MPD and MNTD were created by randomly selecting the observed number of genera in each grid cell 999 times, with all genera in the phylogeny as a sampling pool. Positive values of NRI and NTI indicate phylogenetic clustering, whereas negative values indicate phylogenetic overdispersion in a grid cell. NRI and NTI for woody and herbaceous genera were calculated separately to compare their phylogenetic structures across China.

**Regression analyses between MDT and two climatic variables.** To explore the underlying mechanisms of spatial divergence patterns of the Chinese angiosperms, MDT in each grid cell was regressed against the respective mean values of MAP and MAT in each grid cell using the linear regression model in R. The adjusted  $R^2$  was used to indicate the explanatory power of each variable, although it is clear that these associations do not necessarily indicate causation of the climatic variables in determining MDT. Climatic data were downloaded from the WorldClim database Version 1.4 (<http://www.worldclim.org/>) with a spatial resolution of 10 min<sup>50</sup>.

**Species tree reconstruction and conservation implications.** With our dated genus-level chronogram as the backbone, a species-level tree including 26,978 Chinese angiosperm species was generated by inserting species that were not sampled in our generic tree within the genera to which they belong using the R package 'S.PhyloMaker'<sup>51</sup>. Our species-level tree included approximately 96% of all known angiosperm species native to China; 1,098 aquatic species were not sampled. To mitigate the effect of polytomies on the calculation of phylogenetic diversity, we resolved polytomies in the reconstructed tree with a birth–death clock model<sup>52</sup>. We constructed constraints based on the tree constructed with molecular data, and unresolved taxa were then placed within the relevant constraints. Node heights for each constraint were fixed on the basis of divergence time estimates. We then conducted a Bayesian analysis using MrBayes v.3.2.5<sup>53</sup> with the topological and node height constraints and with the birth–death (speciation and extinction) priors as uniform (0.0, 10.0). Two analyses were run for 2,500,000 generations, sampling every 500 generations, to ensure convergence and mixing; the first 750,000 generations were discarded as burn-in, and 1,000 of the post-burn-in trees were retained for further analyses. The species-level phylogenetic diversity and SES-PD were calculated on the basis of 10 trees randomly selected from the 1,000 trees. The Spearman's rank correlation was used to assess the consistency of phylogenetic diversity or SES-PD patterns based on different trees. Grid cells with the top 5% highest values of both phylogenetic diversity and SES-PD were identified as hotspots of phylogenetic diversity (Fig. 3b). MDT analyses were not conducted on the species tree as the missing data rendered the variation between

replicates uninformative. Once additional molecular information is collected for these species, further analyses can be performed.

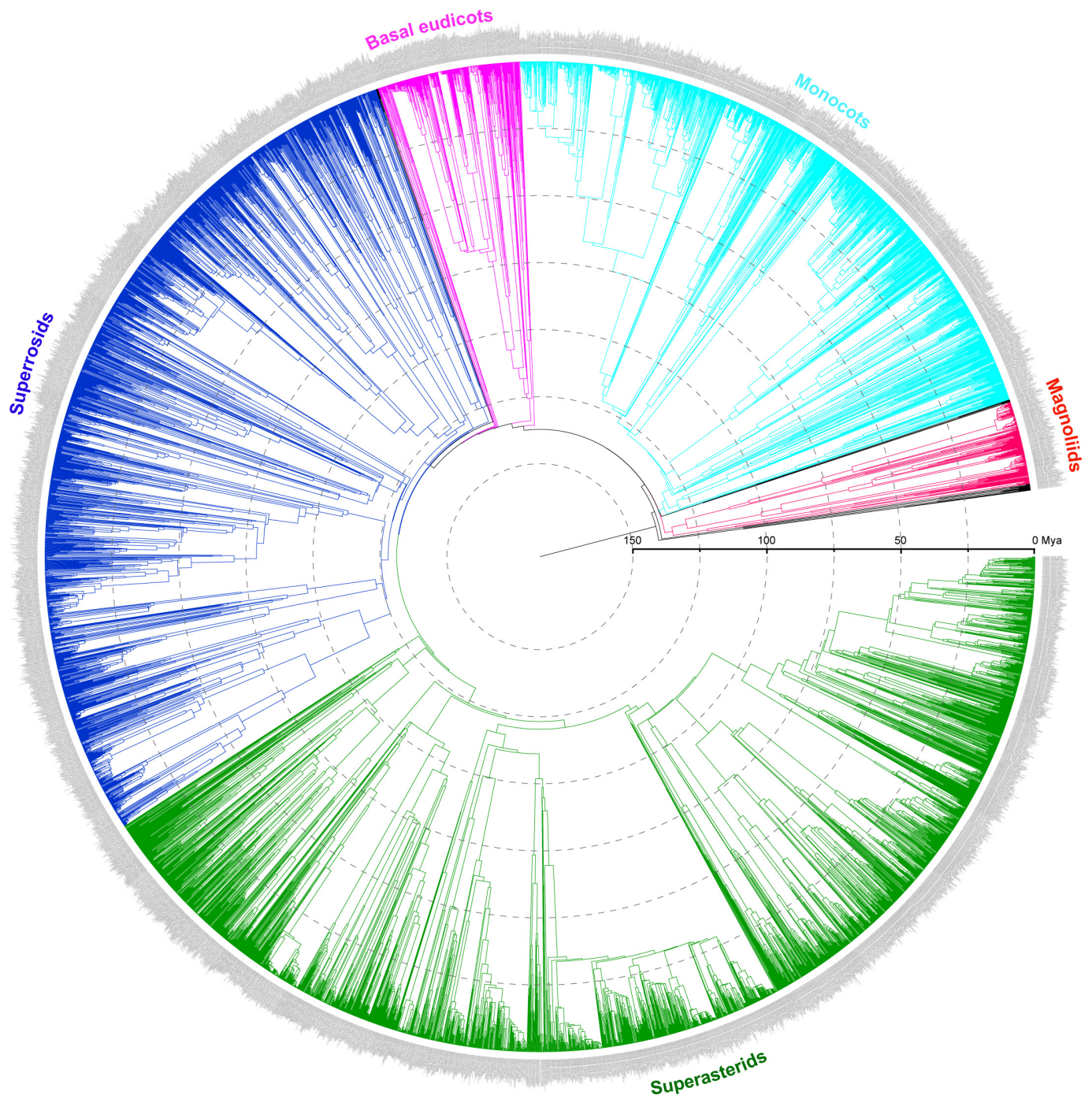
Spatial data of protected areas in China were compiled from two sources: (i) a previous publication<sup>30</sup> that digitized nature reserves in mainland China, which included 334 national, 857 provincial and 1,431 prefectural or county-level nature reserves (provided by Z.-Y. Tang); and (ii) 92 protected areas in Taiwan, downloaded from the Database of Protected Areas (<https://www.protectedplanet.net/>; accessed August 2017). Considering that most of the nature reserves were designed according to administrative units, we calculated richness and phylogenetic diversity in the protected areas with 'county' as the basic unit rather than dividing China into grid cells. Each conservation area was intersected with the map of China to produce the protected areas in ArcGIS. Species occurring in these counties are supposed to be protected, but counties with protected areas that covered less than 10% of the area of a county were excluded to reduce sampling bias.

**Statistics and reproducibility.** No statistical methods were used to predetermine sample size. Spearman's rank correlation and linear regression analyses were conducted in R. Precise  $P$  values are provided to show statistical significance. Null-model tests (999 random replicates) were used to assess the significance of spatial diversity and divergence distributions with  $-1.96$  and  $1.96$  as significant boundaries.

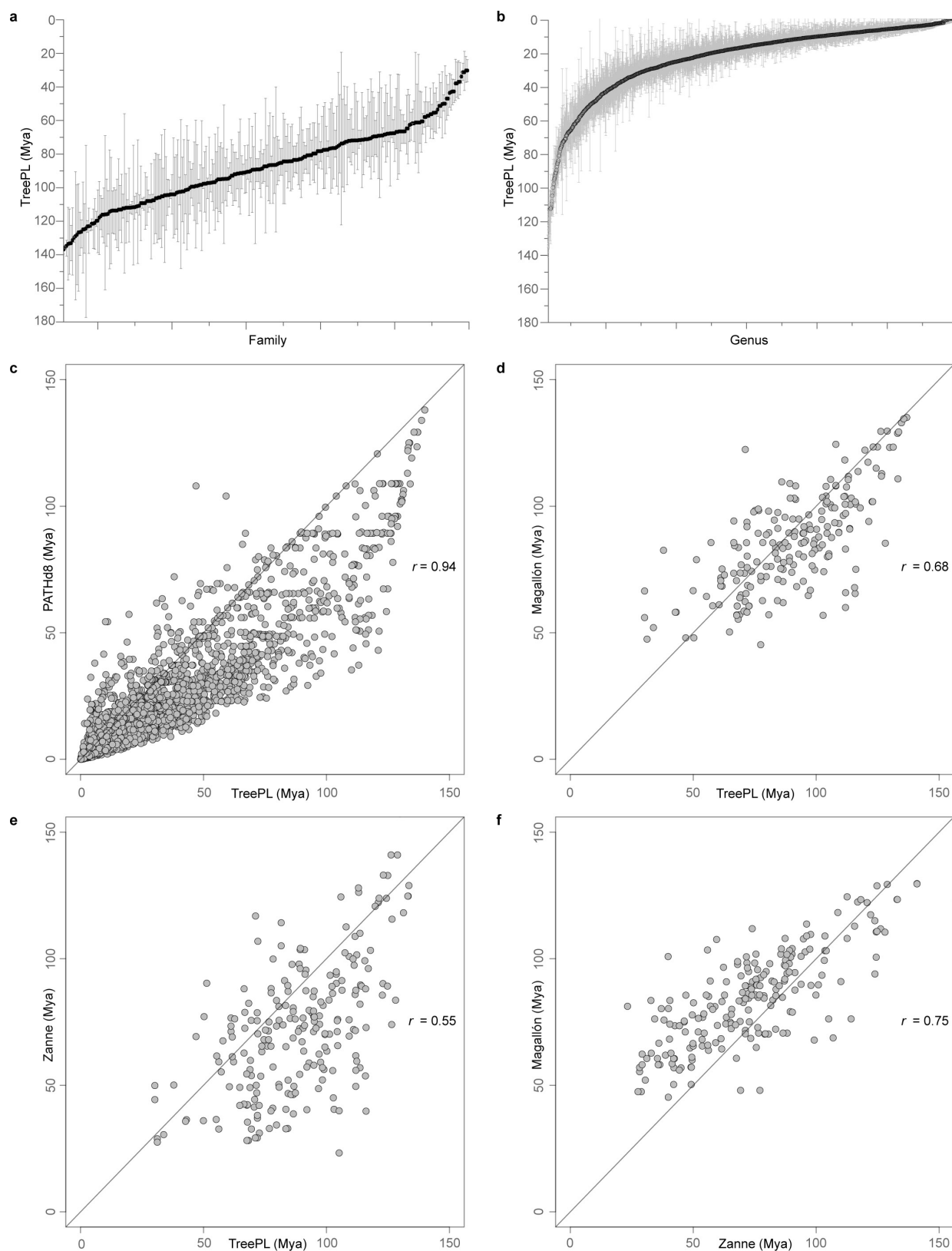
**Code availability.** Example code used to conduct null-model test (written in R) can be found at Dryad: <http://datadryad.org/resource/doi:10.5061/dryad.p89m3>.

**Data availability.** Sequences for phylogenetic analyses have previously been published<sup>31</sup> and deposited in GenBank. The dated phylogeny is archived in Dryad: <http://datadryad.org/resource/doi:10.5061/dryad.p89m3>. The spatial distribution data are available from: [http://www.darwintree.cn/resource/spatial\\_data](http://www.darwintree.cn/resource/spatial_data). All other additional data are available from the corresponding author upon reasonable request.

- Chen, Z.-D. *et al.* Tree of life for the genera of Chinese vascular plants. *J. Syst. Evol.* **54**, 277–306 (2016).
- Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop, IEEE. <http://ieeexplore.ieee.org/document/5676129/> (2010).
- Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
- Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S. & Bremer, K. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **56**, 741–752 (2007).
- Anderson, C. L. *Dating Divergence Times in Phylogenies*. PhD thesis, Uppsala Univ. (2007).
- R Core Team. R: a language and environment for statistical computing. <http://R-project.org/> (2014).
- Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
- Smith, S. A. & Donoghue, M. J. Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**, 86–89 (2008).
- Jetz, W. & Rahbek, C. Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551 (2002).
- Rahbek, C. *et al.* Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proc. R. Soc. Lond. B* **274**, 165–174 (2007).
- Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. Contribution of rarity and commonness to patterns of species richness. *Ecol. Lett.* **7**, 81–87 (2004).
- Iglewicz, B. & Hoaglin, D. *How to Detect and Handle Outliers* (ASQC Quality, 1993).
- Rousseeuw, P. J. & Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **88**, 1273–1283 (1993).
- Cleophas, T. J. Clinical trials: robust tests are wonderful for imperfect data. *Am. J. Ther.* **22**, e1–e5 (2015).
- Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
- Rodrigues, A. S. L., Brooks, T. M. & Gaston, K. J. in *Phylogeny and Conservation* (eds Purvis, A., Gittleman, J. L., & Brooks, T.) 101–119 (Cambridge Univ. Press, 2005).
- Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Qian, H. & Jin, Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* **9**, 233–239 (2016).
- Kuhn, T. S., Mooers, A. Ø. & Thomas, G. H. A simple polytomy resolver for dated phylogenies. *Methods Ecol. Evol.* **2**, 427–436 (2011).
- Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

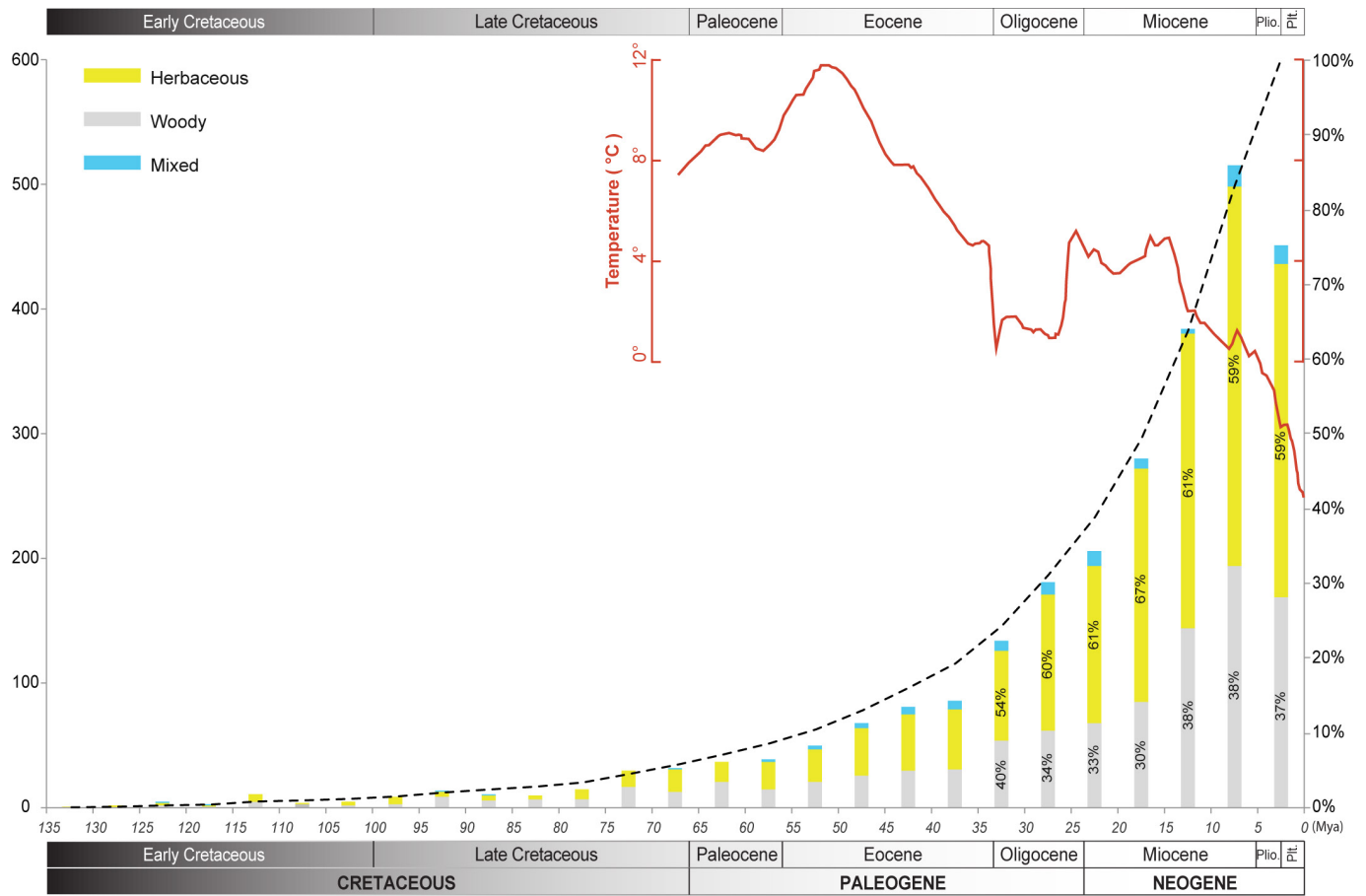


**Extended Data Figure 1 | Dated megaphylogeny of the Chinese angiosperms.** Major clades, including magnoliids, monocots, superrosids and superasterids, as well as the basal eudicot grade, are indicated with different colours. Divergence times were estimated using treePL.



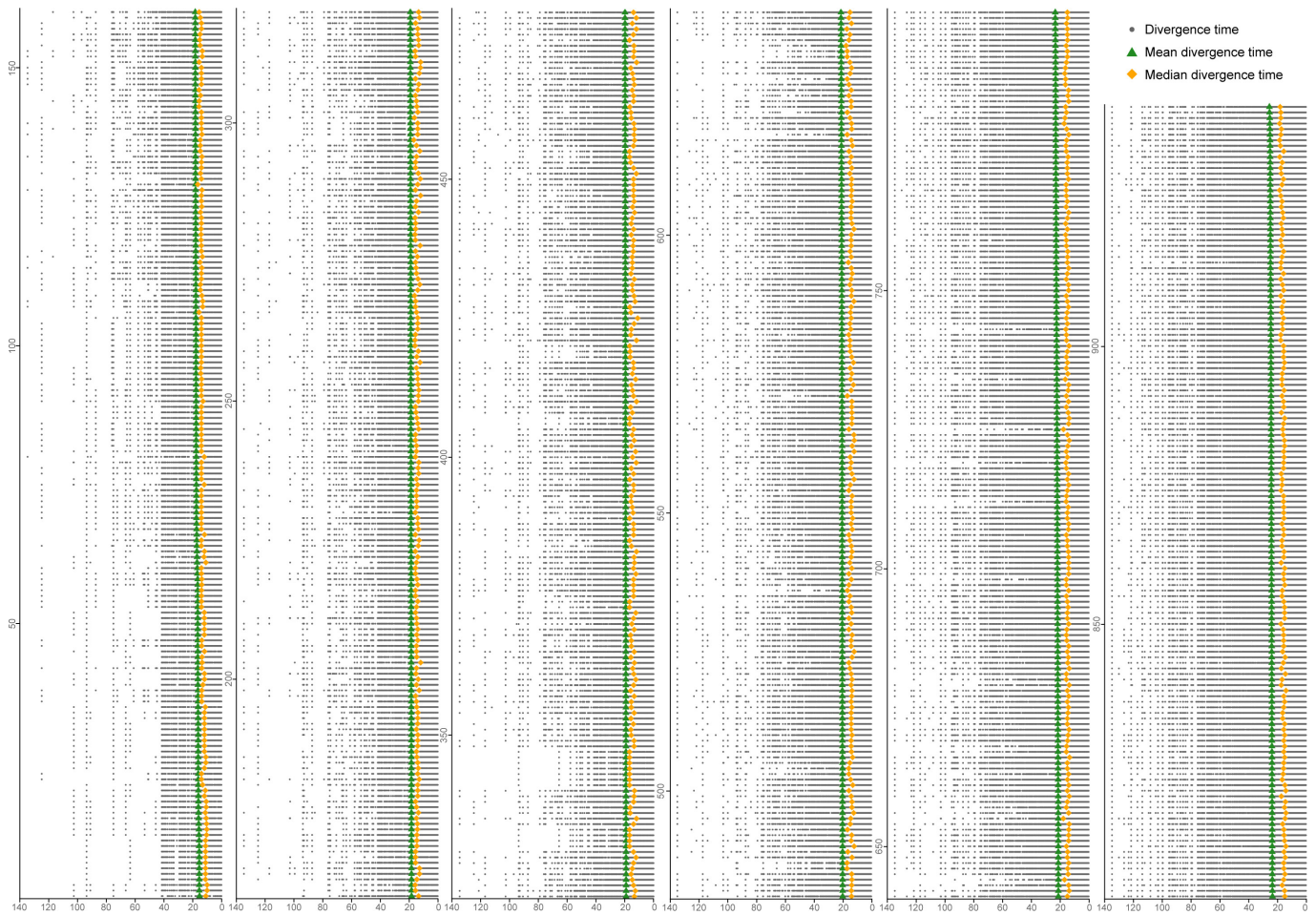
**Extended Data Figure 2 | The 95% confidence intervals of divergence times and the Spearman's rank correlation between our dating and those of recent publications. a, b**, Plots of divergence times and 95% confidence intervals (grey bars) for each family (**a**,  $n = 273$ ) and genus (**b**,  $n = 2,909$ ). The centre values are ages calculated based on the optimal maximum likelihood tree. **c**, Correlation of nodal ages between treePL and

PATHd8 in this study ( $n = 5,863$ ;  $r = 0.94$ ,  $P = 0$ ). **d**, Correlation of family ages between treePL and ref. 22 ( $n = 236$ ,  $r = 0.68$ ,  $P = 1.17 \times 10^{-33}$ ). **e**, Correlation of family ages between treePL and ref. 23 ( $n = 257$ ;  $r = 0.55$ ,  $P = 4.54 \times 10^{-22}$ ). **f**, Correlation of family ages between ref. 22 and ref. 23 ( $n = 235$ ;  $r = 0.75$ ,  $P = 2.11 \times 10^{-43}$ ). The solid line is  $y = x$ .

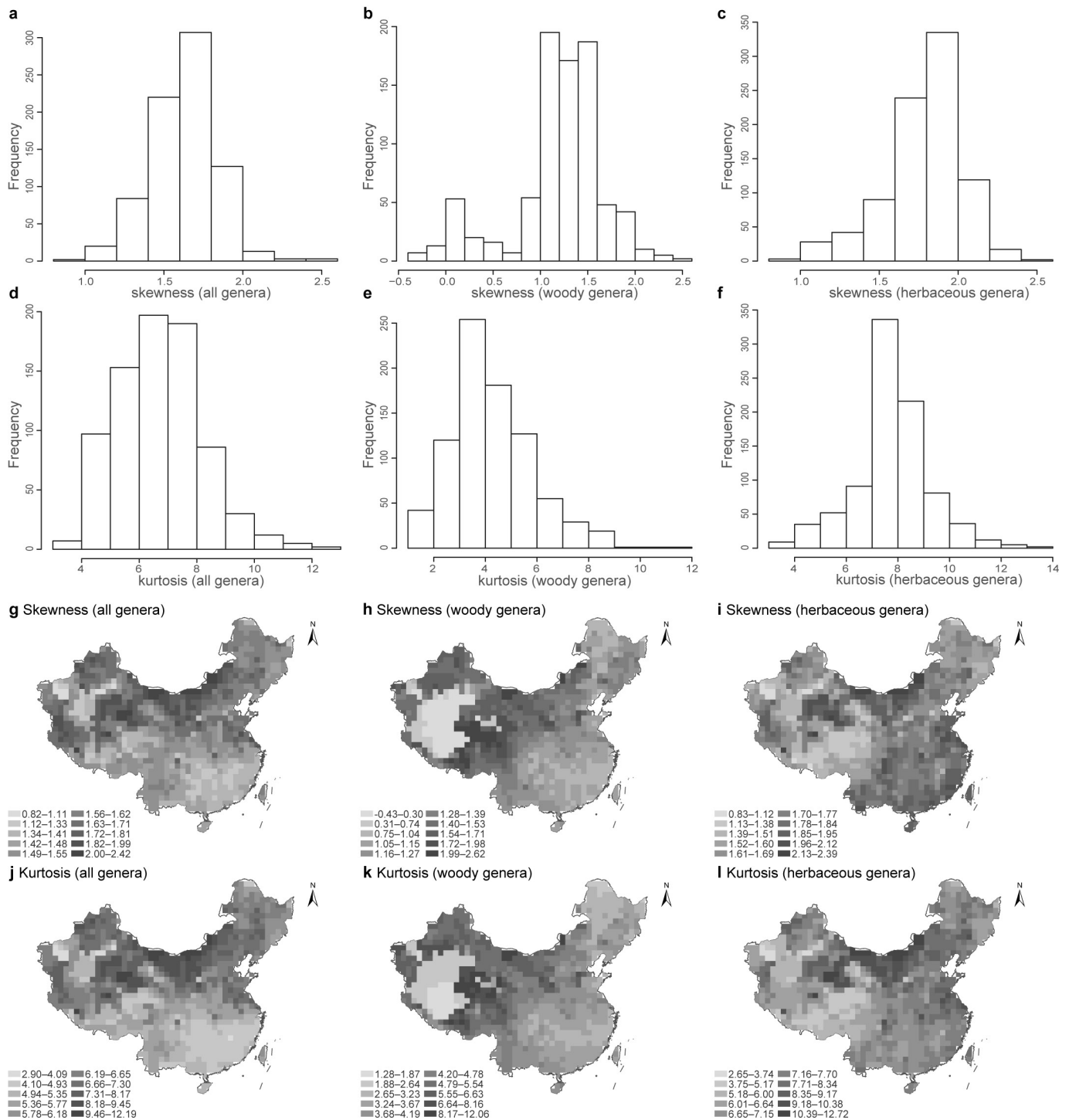


**Extended Data Figure 3 | Number of angiosperm genera that originated during specified geological timespans.** Column with three colours shows the number of woody (grey), herbaceous (yellow) and mixed genera (light blue) that originated within a specific geological timespan. Number of woody genera,  $n = 995$ ; number of herbaceous genera,  $n = 1,569$ ; mixed genera (genera with both woody and herbaceous species),  $n = 101$ . The dashed line indicates the accumulated percentage of genera that have

originated since the Early Cretaceous. Global temperature changes that have occurred since the Palaeogene are shown by the red curve (from ref. 39; reprinted with permission from AAAS). The  $x$  axis indicates the geological period and time in millions of years. The left  $y$  axis shows the total number of genera that have originated by any given time period; the right  $y$ -axis represents the accumulated percentage of genera that originated within a geological time period.

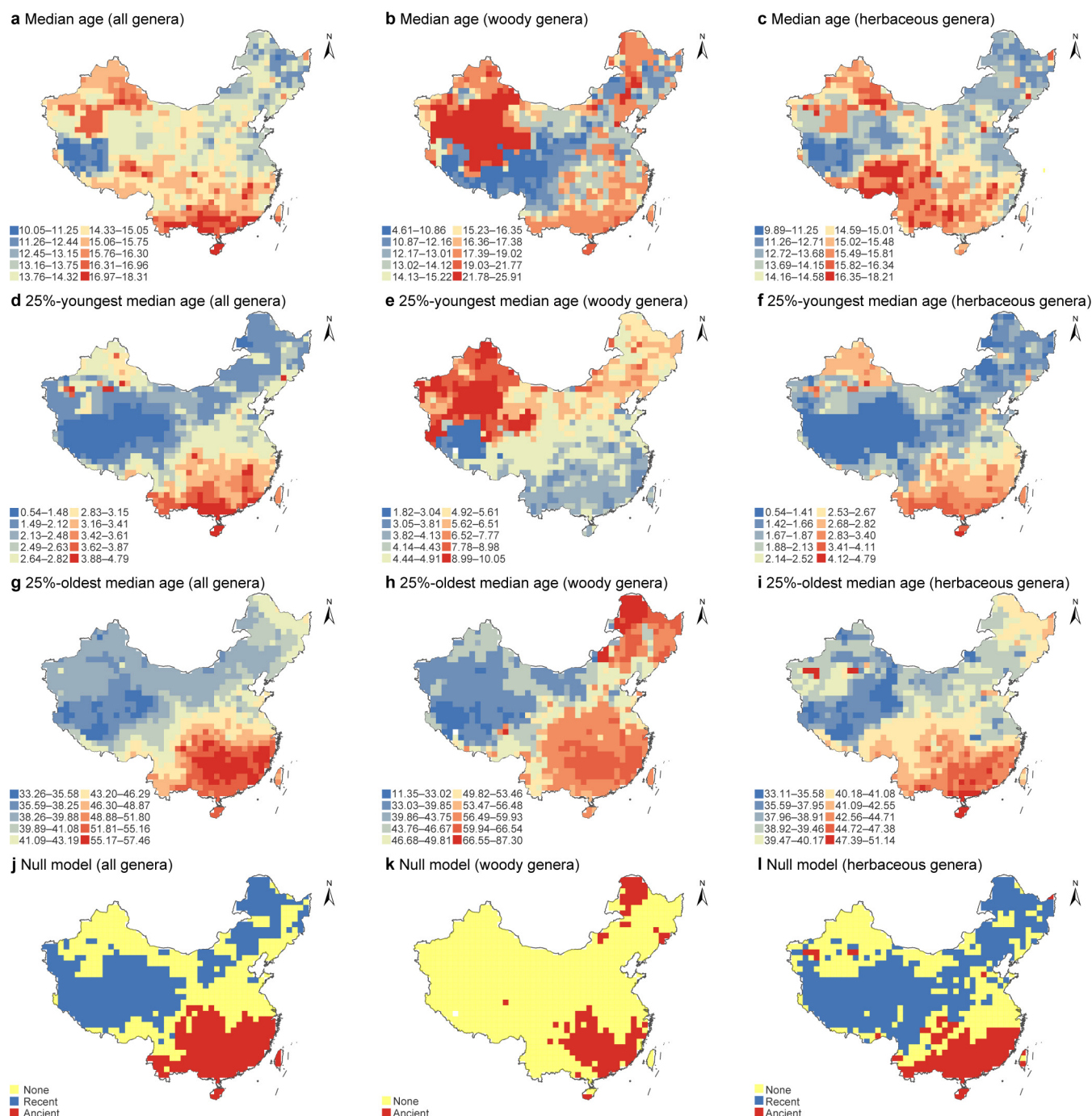


**Extended Data Figure 4 | Plot of divergence times of the Chinese angiosperm genera in each grid cell. Mean and median values of the divergence times are indicated.**



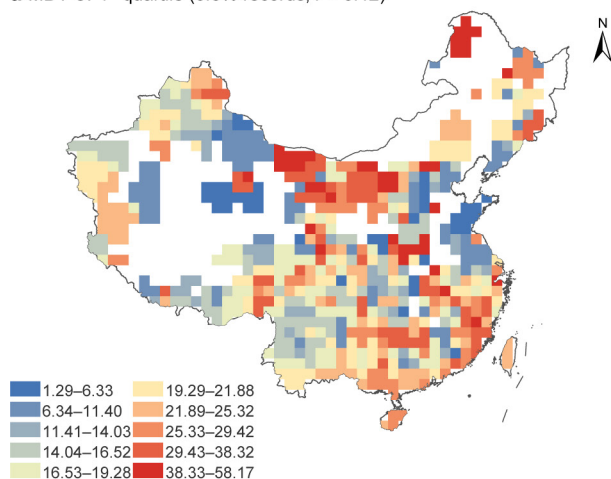
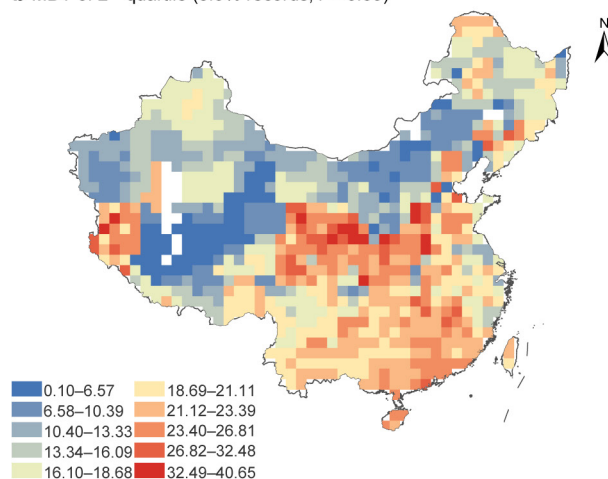
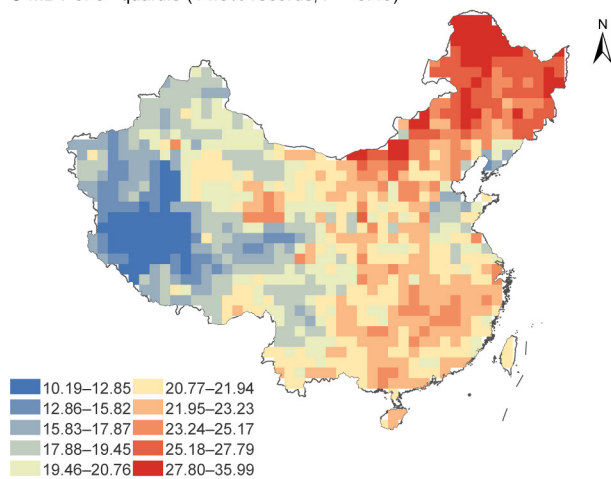
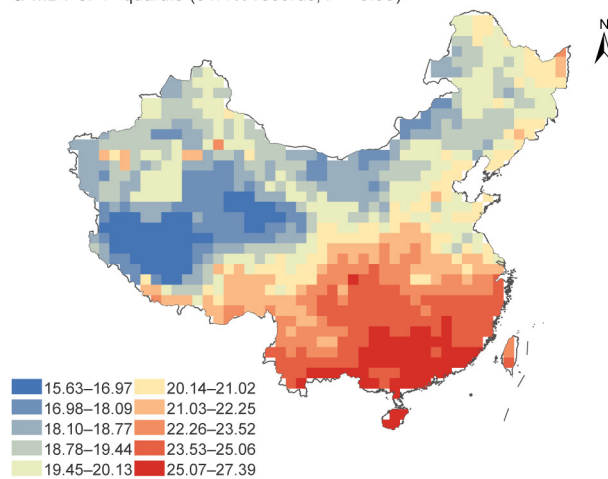
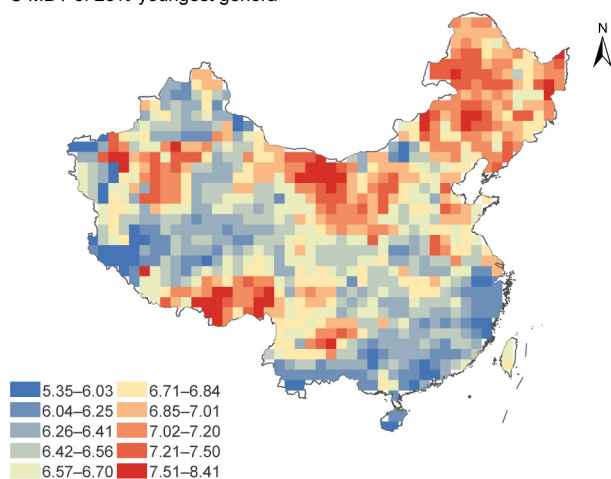
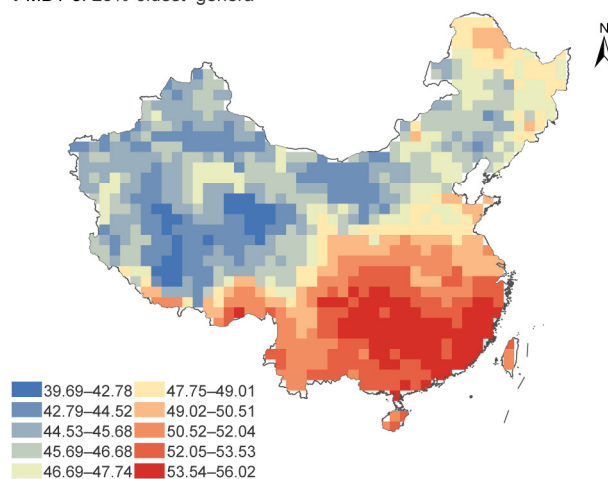
**Extended Data Figure 5 | Histograms and distribution of skewness and kurtosis for divergence times in each grid cell.** **a–c**, Range of skewness for all genera (**a**), woody genera (**b**) and herbaceous genera (**c**). **d–f**, Range of kurtosis (computed as the fourth standardized moment) for all genera (**d**), woody genera (**e**) and herbaceous genera (**f**). **g–i**, Spatial distribution of skewness for all genera (**g**), woody genera (**h**) and herbaceous genera (**i**). **j–l**, Spatial distribution of kurtosis for all genera (**j**), woody genera (**k**) and herbaceous genera (**l**). Skewness values in most grid cells are positive and

around 1–2, which implies that divergence times of genera are slightly right-skewed (there are more young ages in each grid cell). Kurtosis values in most grid cells are within a range of 4–8, larger than the value (3) for a normal distribution, which implies that the distribution of divergence times has more extreme outliers than the normal distribution. For eastern China, kurtosis values of approximately 4 for all genera are consistent with grid cells having a range of divergence times—including very young and very old ages—as expected for an area that is both a cradle and a museum.



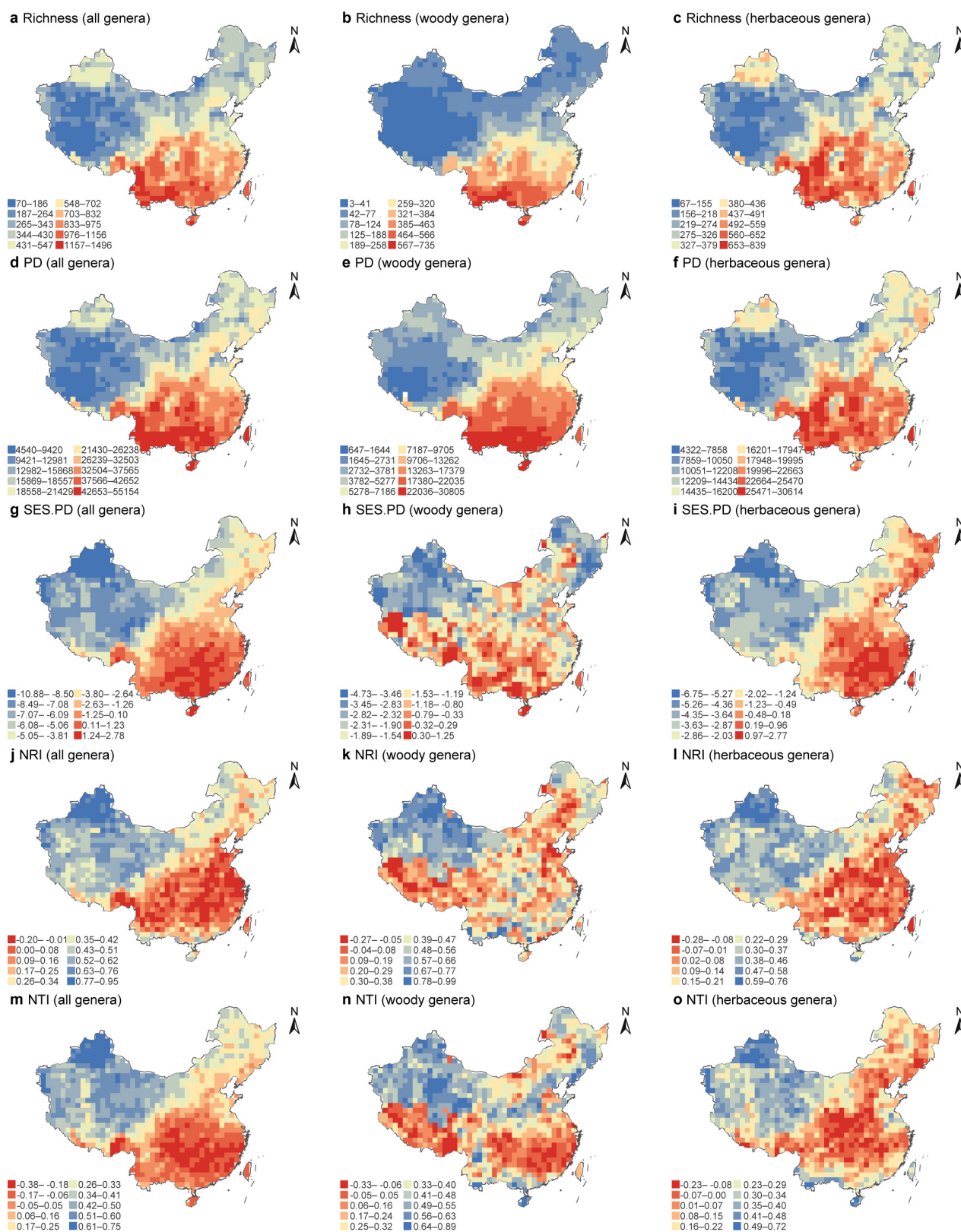
**Extended Data Figure 6 | Geographic patterns of median ages for the Chinese angiosperm genera.** **a–i**, Median ages for all genera, woody genera and herbaceous genera (from left to right), based on all sampled genera (**a–c**), the youngest 25% of genera (**d–f**), and the oldest 25% of genera (**g–i**) in each grid cell. **j–l**, Null-model test to identify recent (blue grid cells) and ancient (red grid cells) divergence centres for all genera (**j**),

woody genera (**k**) and herbaceous genera (**l**). The analyses include 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

**a** MDT of 1<sup>st</sup> quartile (0.8% records,  $r = 0.12$ )**b** MDT of 2<sup>nd</sup> quartile (3.9% records,  $r = 0.59$ )**c** MDT of 3<sup>rd</sup> quartile (14.3% records,  $r = 0.43$ )**d** MDT of 4<sup>th</sup> quartile (81.1% records,  $r = 0.99$ )**e** MDT of 25%-youngest genera**f** MDT of 25%-oldest genera

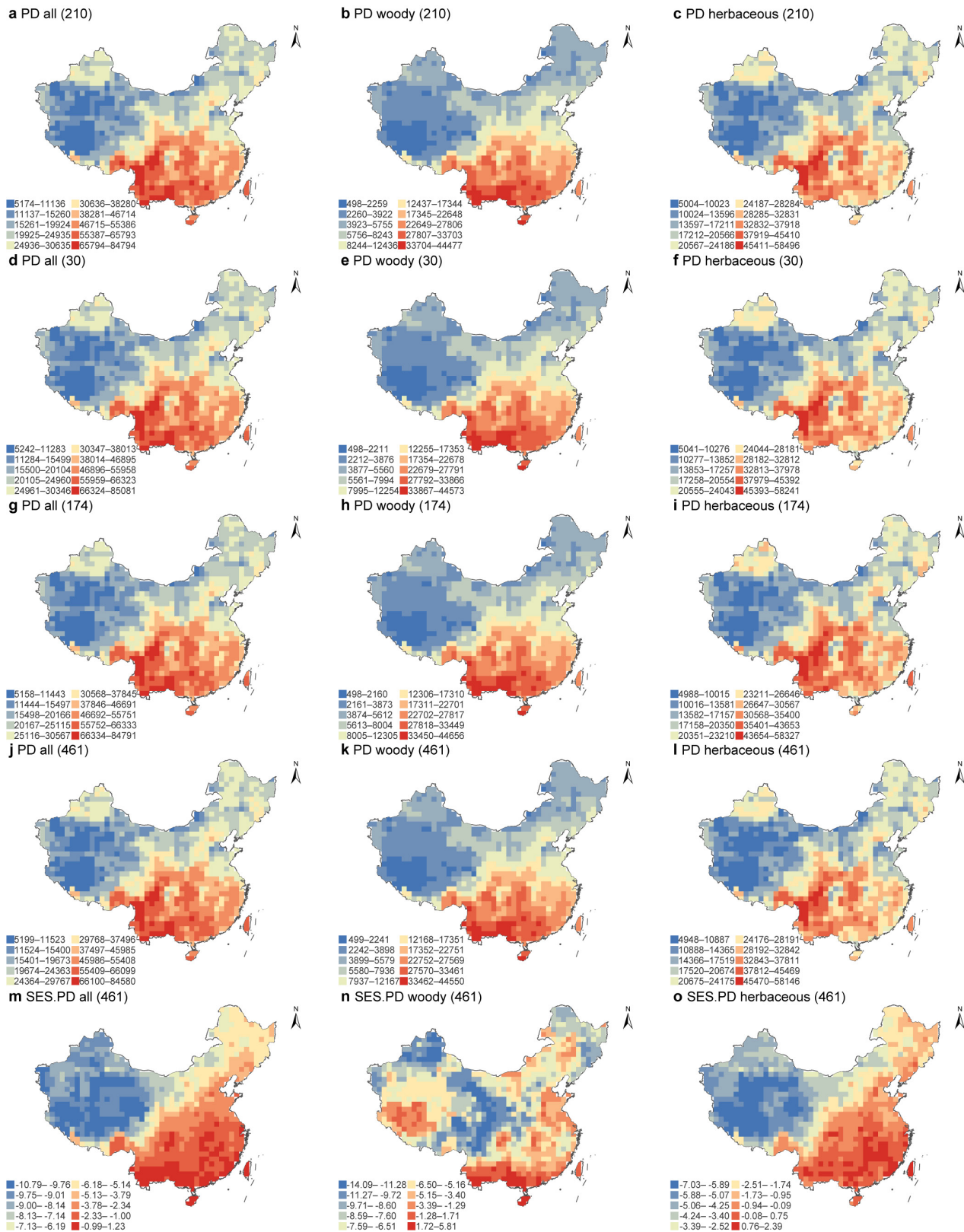
**Extended Data Figure 7 | Spatial distribution of MDTs based on geographic range-size quartiles and the youngest 25% and oldest 25% of genera in China.** **a–d**, MDT patterns of the first (a), second (b), third (c) and fourth quartiles (d) of the sampled Chinese angiosperm genera. The first, second, third and fourth quartiles range from the narrowest to the widest geographic distribution, and represent 0.6%, 3.5%, 13.7% and 82.1% of 1,409,239 records, respectively. The Spearman's rank correlation coefficients between the overall MDT (including all genera)

and MDT of the first, second, third and fourth geographic quartile are 0.12 ( $P = 1.46 \times 10^{-3}$ ), 0.59 ( $P = 1.21 \times 10^{-87}$ ), 0.43 ( $P = 2.51 \times 10^{-43}$ ) and 0.99 ( $P = 0$ ), respectively. **e**, MDT pattern of the youngest 25% of genera in China, showing that there are young genera in both western and eastern China. **f**, MDT pattern of the oldest 25% of genera in China, confirming that older genera mainly occur in eastern China. Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).



**Extended Data Figure 8 | Patterns of generic richness, phylogenetic diversity and phylogenetic structure for the Chinese angiosperm genera.** **a–c**, Richness for all genera (**a**), woody genera (**b**) and herbaceous genera (**c**). **d–f**, Phylogenetic diversity for all genera (**d**), woody genera (**e**) and herbaceous genera (**f**). **g–i**, SES-PD for all genera (**g**), woody genera (**h**) and herbaceous genera (**i**). **j–l**, NRI for all genera (**j**), woody genera (**k**) and

herbaceous genera (**l**). **m–o**, NTI for all genera (**m**), woody genera (**n**) and herbaceous genera (**o**). The analyses include 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).



**Extended Data Figure 9 | Patterns of species-level phylogenetic diversity for all Chinese angiosperms. a–l,** Observed phylogenetic diversity for all species (a, d, g, j), woody species (b, e, h, k) and herbaceous species (c, f, i, l) based on species trees 210, 30, 174 and 461 (species trees were randomly selected from 1,000 post-burn-in trees). **m–o,** SES-PD for all species (m), woody species (n) and herbaceous species (o) based on

species tree 461. The analyses include 26,978 angiosperm species (woody,  $n = 10,169$ ; herbaceous,  $n = 16,809$ ). Phylogenetic diversity and SES-PD based on 10 species trees produce similar patterns; Spearman's rank correlation coefficients,  $r > 0.99$ ,  $P < 2.20 \times 10^{-16}$ . Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

**Extended Data Table 1 | Number of genera that occur only in western or eastern China, with the number of woody, herbaceous and mixed genera in each order indicated**

|               | Rank | Order          | No. of genera | Woody | Herbaceous | Mixed |
|---------------|------|----------------|---------------|-------|------------|-------|
| Western China | 1    | Asterales      | 26            | 0     | 26         | 0     |
|               | 2    | Brassicales    | 23            | 0     | 23         | 0     |
|               | 3    | Caryophyllales | 18            | 3     | 14         | 1     |
|               | 4    | Apiales        | 10            | 0     | 10         | 0     |
|               | 5    | Fabales        | 8             | 4     | 3          | 1     |
|               | 6    | Poales         | 6             | 2     | 4          | 0     |
|               | 7    | Lamiales       | 5             | 0     | 5          | 0     |
|               | 8    | Malvales       | 3             | 0     | 3          | 0     |
|               | 9    | Asparagales    | 3             | 0     | 3          | 0     |
|               | 10   | Boraginales    | 2             | 0     | 2          | 0     |
|               | 11   | Alismatales    | 1             | 0     | 1          | 0     |
|               | 12   | Ranunculales   | 1             | 0     | 1          | 0     |
|               | 13   | Malpighiales   | 1             | 0     | 1          | 0     |
|               | 14   | Myrtales       | 1             | 0     | 1          | 0     |
|               | 15   | Rosales        | 1             | 1     | 0          | 0     |
|               | 16   | Saxifragales   | 1             | 0     | 1          | 0     |
|               | 17   | Zygophyllales  | 1             | 1     | 0          | 0     |
| Eastern China | 1    | Lamiales       | 112           | 25    | 72         | 15    |
|               | 2    | Gentianales    | 102           | 82    | 18         | 2     |
|               | 3    | Asparagales    | 98            | 1     | 96         | 1     |
|               | 4    | Poales         | 81            | 16    | 65         | 0     |
|               | 5    | Malpighiales   | 72            | 60    | 9          | 3     |
|               | 6    | Fabales        | 54            | 37    | 16         | 1     |
|               | 7    | Sapindales     | 46            | 43    | 3          | 0     |
|               | 8    | Asterales      | 45            | 7     | 35         | 3     |
|               | 9    | Ericales       | 32            | 27    | 5          | 0     |
|               | 10   | Alismatales    | 30            | 0     | 30         | 0     |
|               | 11   | Malvales       | 29            | 24    | 3          | 2     |
|               | 12   | Myrtales       | 29            | 24    | 3          | 2     |
|               | 13   | Magnoliales    | 25            | 25    | 0          | 0     |
|               | 14   | Rosales        | 24            | 20    | 4          | 0     |
|               | 15   | Ranunculales   | 23            | 14    | 9          | 0     |
|               | 16   | Saxifragales   | 19            | 14    | 5          | 0     |
|               | 17   | Caryophyllales | 18            | 3     | 14         | 1     |
|               | 18   | Solanales      | 17            | 8     | 6          | 3     |
|               | 19   | Zingiberales   | 16            | 1     | 15         | 0     |
|               | 20   | Apiales        | 15            | 8     | 6          | 1     |
|               | 21   | Cucurbitales   | 13            | 2     | 11         | 0     |
|               | 22   | Laurales       | 13            | 12    | 1          | 0     |
|               | 23   | Arecales       | 13            | 13    | 0          | 0     |
|               | 24   | Santalales     | 12            | 12    | 0          | 0     |
|               | 25   | Cornales       | 10            | 7     | 3          | 0     |
|               | 26   | Brassicales    | 9             | 5     | 4          | 0     |
|               | 27   | Celastrales    | 9             | 9     | 0          | 0     |
|               | 28   | Icacinales     | 7             | 7     | 0          | 0     |
|               | 29   | Oxalidales     | 6             | 5     | 1          | 0     |
|               | 30   | Fagales        | 6             | 6     | 0          | 0     |
|               | 31   | Commelinales   | 6             | 0     | 6          | 0     |
|               | 32   | Boraginales    | 5             | 1     | 4          | 0     |
|               | 33   | Liliales       | 4             | 0     | 4          | 0     |
|               | 34   | Aquifoliales   | 3             | 2     | 1          | 0     |
|               | 35   | Pandanales     | 3             | 1     | 2          | 0     |
|               | 36   | Dipsacales     | 3             | 3     | 0          | 0     |
|               | 37   | Piperales      | 3             | 0     | 3          | 0     |
|               | 38   | Dioscoreales   | 2             | 0     | 2          | 0     |
|               | 39   | Dilleniales    | 2             | 2     | 0          | 0     |
|               | 40   | Chloranthales  | 2             | 0     | 1          | 1     |
|               | 41   | Huerteales     | 2             | 2     | 0          | 0     |
|               | 42   | Nymphaeales    | 1             | 0     | 1          | 0     |
|               | 43   | Proteales      | 1             | 1     | 0          | 0     |
|               | 44   | Petrosaviales  | 1             | 0     | 1          | 0     |
|               | 45   | Metteniusales  | 1             | 1     | 0          | 0     |
|               | 46   | Escalloniales  | 1             | 1     | 0          | 0     |
|               | 47   | Vitales        | 1             | 1     | 0          | 0     |

Mixed, genera with both woody and herbaceous species.

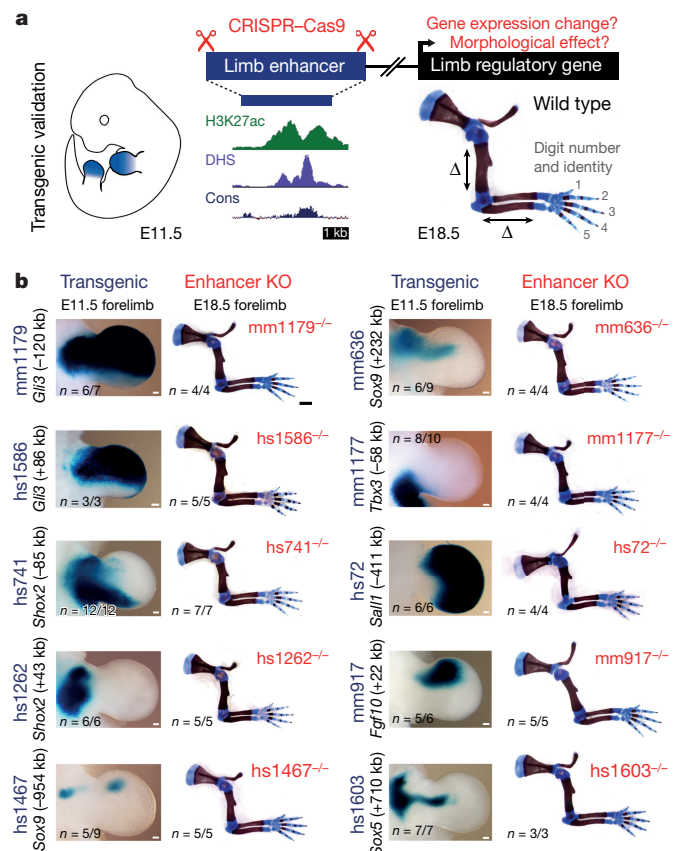
# Enhancer redundancy provides phenotypic robustness in mammalian development

Marco Osterwalder<sup>1</sup>, Iros Barozzi<sup>1</sup>, Virginie Tissi res<sup>2,3</sup>, Yoko Fukuda-Yuzawa<sup>1</sup>, Brandon J. Mannion<sup>1</sup>, Sarah Y. Afzal<sup>1</sup>, Elizabeth A. Lee<sup>1</sup>, Yiwen Zhu<sup>1</sup>, Ingrid Plajzer-Frick<sup>1</sup>, Catherine S. Pickle<sup>1</sup>, Momoe Kato<sup>1</sup>, Tyler H. Garvin<sup>1</sup>, Quan T. Pham<sup>1</sup>, Anne N. Harrington<sup>1</sup>, Jennifer A. Akiyama<sup>1</sup>, Veena Afzal<sup>1</sup>, Javier Lopez-Rios<sup>2,3</sup>, Diane E. Dickel<sup>1</sup>, Axel Visel<sup>1,4,5</sup> & Len A. Pennacchio<sup>1,4,6</sup>

Distant-acting tissue-specific enhancers, which regulate gene expression, vastly outnumber protein-coding genes in mammalian genomes, but the functional importance of this regulatory complexity remains unclear<sup>1,2</sup>. Here we show that the pervasive presence of multiple enhancers with similar activities near the same gene confers phenotypic robustness to loss-of-function mutations in individual enhancers. We used genome editing to create 23 mouse deletion lines and inter-crosses, including both single and combinatorial enhancer deletions at seven distinct loci required for limb development. Unexpectedly, none of the ten deletions of individual enhancers caused noticeable changes in limb morphology. By contrast, the removal of pairs of limb enhancers near the same gene resulted in discernible phenotypes, indicating that enhancers function redundantly in establishing normal morphology. In a genetic background sensitized by reduced baseline expression of the target gene, even single enhancer deletions caused limb abnormalities, suggesting that functional redundancy is conferred by additive effects of enhancers on gene expression levels. A genome-wide analysis integrating epigenomic and transcriptomic data from 29 developmental mouse tissues revealed that mammalian genes are very commonly associated with multiple enhancers that have similar spatiotemporal activity. Systematic exploration of three representative developmental structures (limb, brain and heart) uncovered more than one thousand cases in which five or more enhancers with redundant activity patterns were found near the same gene. Together, our data indicate that enhancer redundancy is a remarkably widespread feature of mammalian genomes that provides an effective regulatory buffer to prevent deleterious phenotypic consequences upon the loss of individual enhancers.

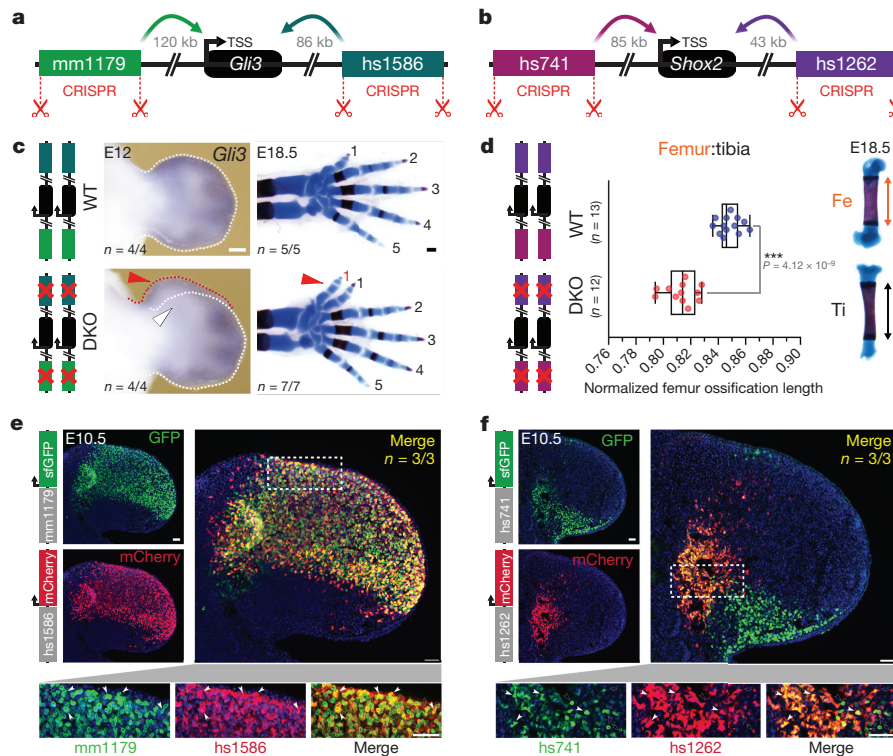
Enhancers are a principal class of *cis*-regulatory elements that orchestrate precise gene expression patterns, which are essential for numerous processes including embryonic development<sup>2</sup>. They are now routinely predicted by genome-wide chromatin profiling methods, which identify positions of open chromatin or enhancer-associated histone marks<sup>3</sup>. Enhancers predicted by these high-throughput approaches outnumber genes by approximately an order of magnitude<sup>1</sup>, raising the question of their functional significance. In particular, it remains unclear whether mammalian enhancers typically regulate complementary spatiotemporal aspects of gene expression in an additive fashion<sup>4–7</sup>, or if this regulatory complexity more commonly results in functional redundancy among enhancers associated with the same gene<sup>8–10</sup>.

Using the developing limb as a model for gene regulation during morphogenetic processes<sup>11,12</sup>, we investigated the functional importance of enhancers *in vivo*. We used CRISPR–Cas9 genome editing to



**Figure 1 | Lack of limb morphological abnormalities in ten enhancer deletion lines.** **a**, All selected enhancers are active in the limb mesenchyme (blue shading) at E11.5, are marked by epigenomic H3K27 acetylation and DNase I hypersensitivity (DHS) at E11.5, and contain a conserved core sequence (Cons). Target gene expression and limb morphology were assessed following deletion of individual enhancers (Extended Data Fig. 1a–j). **b**, None of the individual enhancer deletions caused obvious defects in the structure of skeletal elements. Enhancers are identified by VISTA ID numbers. Enhancer activities (left, E11.5) and forelimb skeletons of enhancer knockout (KO) embryos (right, E18.5) are shown (see Extended Data Fig. 3 for wild-type controls). Predicted target gene and enhancer distance (+, downstream; –, upstream) from the transcriptional start site (TSS) are indicated. *n* represents independent biological replicates with similar results. Scale bars, 100 μm (E11.5), 1 mm (E18.5).

<sup>1</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. <sup>2</sup>Department of Biomedicine, University of Basel, 4058 Basel, Switzerland. <sup>3</sup>Centro Andaluz de Biología del Desarrollo, CSIC/JA/Universidad Pablo de Olavide, 41013 Seville, Spain. <sup>4</sup>US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. <sup>5</sup>School of Natural Sciences, University of California, Merced, California 95343, USA. <sup>6</sup>Comparative Biochemistry Program, University of California, Berkeley, California 94720, USA.



**Figure 2 | Pairwise loss of limb enhancers with overlapping activities results in morphological abnormalities.** **a, b**, CRISPR-deleted enhancers and their distance to the TSSs of predicted target genes (*Gli3*, *Shox2*). **c**, Left, RNA *in situ* hybridization reveals reduced *Gli3* expression in anterior hand plates of mm1179/hs1586 DKO embryos (white arrowhead). Red arrowhead, local expansion of anterior mesenchyme, a hallmark of *Gli3* deficiency. Right, forelimb skeletons with digits labelled 1 to 5, from anterior to posterior. DKO embryos exhibit duplication of digit 1 (red arrowhead). Scale bars, 200  $\mu$ m. WT, wild type. **d**, Reduced femur ossification length in hs741/hs1262 DKO embryos (normalized to tibia

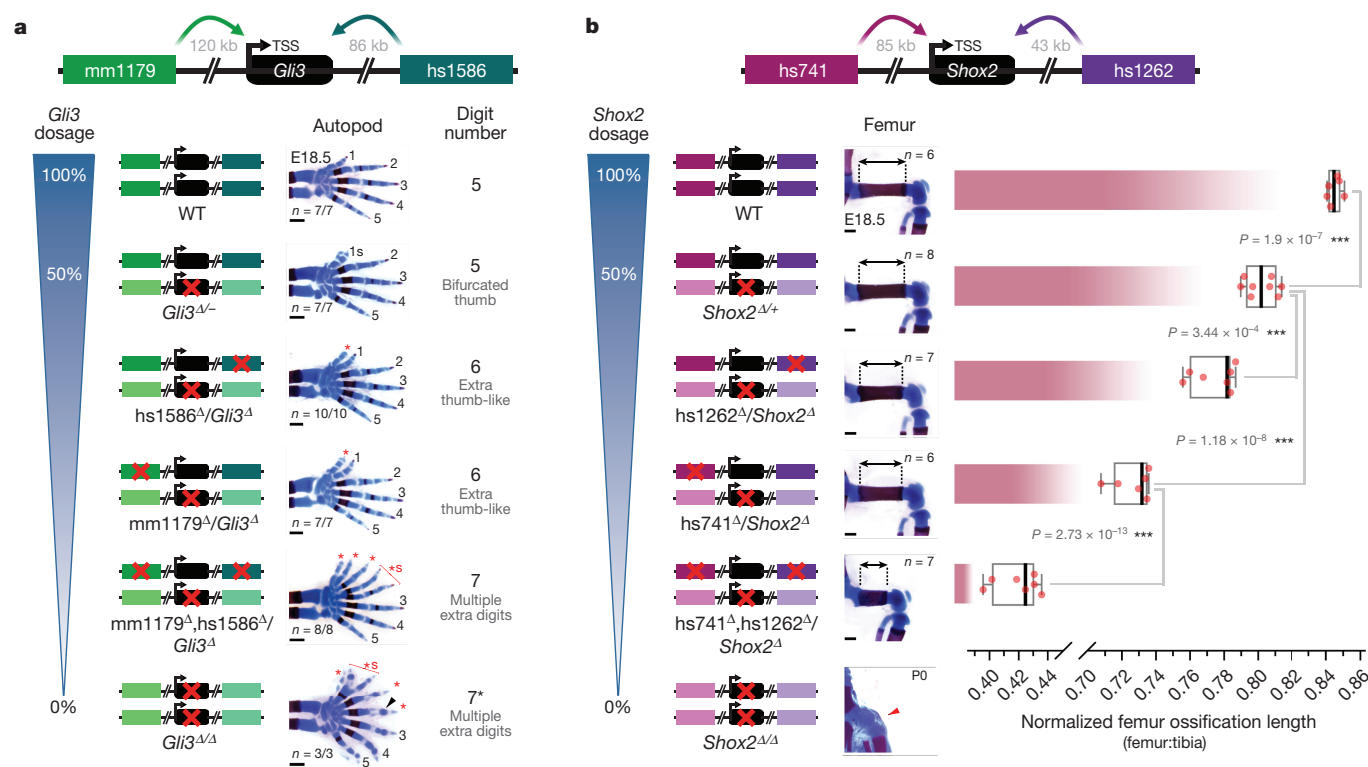
ossification length). Box plot indicates median, interquartile values, range and individual biological replicates. \*\*\* $P < 0.001$  (two-tailed, unpaired *t*-test). **e, f**, Co-localization of *Gli3* (**e**; mm1179, green; hs1586, red) and *Shox2* (**f**; hs741, green; hs1262, red) enhancer activities via enhancer-reporter transgenes and immunofluorescence in forelimb buds of double transgenic embryos. White arrowheads indicate examples of double-positive cells. Empty arrowheads or arrows indicate cells marked by single enhancers. Nuclei are stained blue. Scale bars, 50  $\mu$ m. *n* represents independent biological replicates with similar results.

individually delete ten embryonic enhancers, each with strong evolutionary conservation and robust limb activity in transgenic mouse reporter assays<sup>13–17</sup> (VISTA Enhancer Browser: <https://enhancer.lbl.gov/>) (Fig. 1a, Extended Data Fig. 1a–j and Supplementary Table 1). Each enhancer (identified by VISTA ID number) is located in the vicinity of a gene associated with human congenital limb malformations, and deletion of these genes in mice results in limb phenotypes ranging from polydactyly (*Gli3*) to complete loss of limbs (*Fgf10*) (Extended Data Fig. 1 and Supplementary Table 2). In all cases, the limb activity pattern of the enhancer at embryonic day 11.5 (E11.5) overlaps spatial RNA expression of the associated target gene, suggesting that these enhancers are part of the regulatory architecture that controls the expression of these genes<sup>16–21</sup> (Extended Data Fig. 2). Capture-C chromatin conformation data from embryonic limbs<sup>22</sup> confirmed that at least six of these enhancers physically interacted with their predicted target genes (Extended Data Fig. 1k). This framework enabled us to investigate the functional contribution of each enhancer by comparing the potential limb skeletal abnormalities caused by enhancer loss to the phenotypes observed in gene knockout mice.

Unexpectedly, we did not detect any abnormalities in bone number, shape, length, position or mineralization in mice in which any of the ten single enhancers was deleted (Fig. 1b and Extended Data Fig. 3). Similarly, we observed neither significant differences in predicted target gene expression in embryonic limbs for nine out of ten individual enhancer deletions, nor obvious changes in local H3K27ac (acetylation of lysine 27 on histone H3) signatures outside the deleted enhancers (Extended Data Figs 2, 4). Together, these results suggest that a substantial proportion of limb enhancers, even if highly conserved in evolution, are not individually essential for normal limb morphogenesis.

One possible explanation for the lack of an obvious phenotype in individual limb enhancer knockout lines is that different enhancers associated with the same gene may have spatiotemporally redundant, rather than unique, activity. Our selected panel of enhancers (Fig. 1b and Extended Data Fig. 1a–j) included three enhancer pairs with overlapping limb activity domains and the same predicted target gene (mm1179–hs1586, hs741–hs1262, and hs1467–mm636; Extended Data Fig. 5a–c). Using iterative CRISPR–Cas9 genome editing, we generated double enhancer knockout (DKO) mice for each enhancer pair (Extended Data Fig. 5a–d, g, j), such that both deletions occurred *in cis*. In two out of three cases, involving enhancer pairs near *Gli3* and *Shox2*, homozygous DKO embryos showed phenotypic abnormalities affecting skeletal limb morphology (Fig. 2a–d and Extended Data Fig. 5f, i, j). Mice lacking both enhancers near *Gli3* (mm1179 and hs1586) had substantially reduced *Gli3* expression in the embryonic hand plate and exhibited forelimb-specific polydactyly (Fig. 2c and Extended Data Fig. 5e, f), a phenotypic hallmark of diminished *Gli3* expression<sup>23,24</sup>. In addition, combined deletion of the two enhancers near *Shox2* (hs741 and hs1262) reduced *Shox2* expression, predominantly in embryonic hindlimbs, and resulted in a marked reduction in femur ossification (Fig. 2d and Extended Data Fig. 5h, i), consistent with the stylopod reductions observed when the *Shox2* gene is inactivated<sup>18,25</sup>. Together, these results show that although each of the four enhancers near *Gli3* and *Shox2* is individually dispensable for limb morphology, the respective pairs of enhancers are collectively required for normal limb development.

To examine the degree of overlap between the activity patterns of phenotypically redundant enhancers at the cellular level, we generated transgenic mouse lines expressing fluorescent reporters under the



**Figure 3 | Normally dispensable individual enhancers are required for limb morphology in a sensitized background.** Individual and combined enhancer deletions in the presence of only one copy of the *Gli3* (a) or *Shox2* (b) target genes and the resulting limb morphology at E18.5. Wedges indicate inferred gene dosage. **a**, Skeletal forelimb autopod phenotypes at E18.5 resulting from mm1179 and hs1586 enhancer deletions in the presence of reduced *Gli3* dosage. 1–5, normal digits. Red asterisk, extra digits with unclear identity. \*, 'split' digit. Black asterisk and arrowhead indicate the presence of hypoplastic distal phalanges.

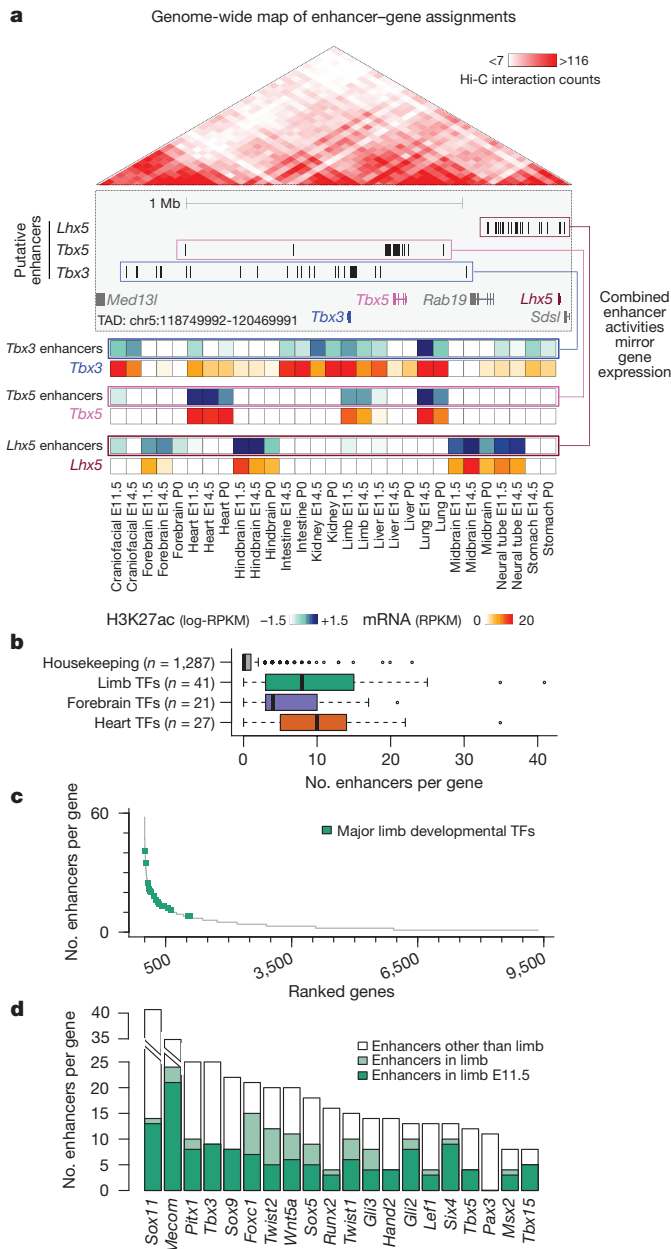
control of each of the *Gli3* or *Shox2* enhancers (mm1179–GFP, hs1586–mCherry, hs741–GFP and hs1262–mCherry). Using immunofluorescence on limb sections from double transgenic embryos, we tracked the activity of each of the four enhancers during limb development (Fig. 2e, f and Extended Data Fig. 6). Consistent with the preaxial polydactyly observed in *Gli3* DKO embryos, limb progenitor cells marked by both *Gli3* enhancers were observed at high density in the anterior limb mesenchyme (Fig. 2e and Extended Data Fig. 6c, d). In *Shox2* double enhancer reporter embryos, a major accumulation of cells with dual *Shox2* enhancer activities is present in a proximal limb mesenchymal cell population known to harbour stylopod progenitors<sup>12</sup> (Fig. 2f). In conjunction with our deletion studies, these results illustrate the degree of functional overlap between pairs of enhancers near the same gene at the cellular level.

Considering the apparent contrast between the morphological redundancy of pairs of enhancers and the strong evolutionary conservation of each individual enhancer, we studied the phenotypic effect of single and combinatorial enhancer deletions in sensitized genetic backgrounds carrying heterozygous deletions of the presumptive target genes (Fig. 3). We used CRISPR–Cas9 to engineer *Gli3* and *Shox2* gene loss-of-function alleles, which recapitulated expected gene dosage reductions and previously published phenotypes (Extended Data Figs 7, 8). We then used these alleles to generate compound heterozygous mice harbouring one or more disrupted enhancers with a wild-type gene on one allele and a disrupted gene but wild-type enhancers on the other allele (Fig. 3). For *Gli3*, the absence of either enhancer (mm1179 or hs1586) in the presence of only one functional *Gli3* allele resulted in a supernumerary anterior digit (Fig. 3a and Extended

**b**, Progressive reduction in femur ossification length (double arrows) due to hs741 and hs1262 enhancer loss in a *Shox2*-sensitized background. The relative length of femur ossification, normalized to the tibia ossification length, is shown. For comparison, the bottom panel shows absence of femur ossification in *Shox2*-deficient limbs at P0 (red arrowhead, reproduced with permission from ref. 25). *n* represents number of independent biological replicates with similar results. Box plots indicate median, interquartile values, range and individual biological replicates. \*\*\* $P < 0.001$  (two-tailed, unpaired *t*-test). Scale bars, 500  $\mu\text{m}$ .

Data Fig. 8a), which is more severe than the terminally bifurcated thumb observed in *Gli3* heterozygotes (Fig. 3a). Similarly, for *Shox2* the removal of either neighbouring enhancer (hs1262 or hs741) in combination with compound heterozygous deletion of the *Shox2* gene resulted in a more pronounced reduction in femur length than observed in *Shox2* heterozygotes (Fig. 3b). For both pairs of enhancers, compound heterozygous mice carrying deletions of both enhancers on one allele and a deletion of the gene on the other allele showed even more severe phenotypes. In the case of *Gli3*, loss of both enhancers over a *Gli3* null allele resulted in greatly reduced expression of *Gli3* (Extended Data Fig. 7b, c) and severe pre-axial polydactyly in forelimbs, similar in severity to homozygous loss of the *Gli3* gene<sup>24</sup> (Fig. 3a and Extended Data Fig. 8a). Likewise, compound heterozygous deletion of enhancers hs741 and hs1262 over a *Shox2* gene deletion strongly reduced *Shox2* expression (Extended Data Fig. 7e, f) and resulted in a severe reduction in femur length and substantial shortening of the humerus (Fig. 3b and Extended Data Fig. 8b, c), consistent with the phenotypes that result from homozygous *Shox2* gene loss<sup>18,25</sup>. Together, our data demonstrate that these developmental enhancers, although seemingly dispensable under non-sensitized conditions, show individual functional contributions to limb development under conditions of reduced genetic robustness.

The lack of phenotypic change upon deletion of individual enhancers, and the functional redundancy observed among enhancer pairs, raises the question of how commonly such redundancy occurs in mammalian gene regulatory landscapes. To explore this question systematically, we devised a genome-wide, correlation-based computational approach to estimate the number of enhancers that regulate each gene



**Figure 4 | Enhancers with redundant signatures are prevalent near developmental genes.** **a**, Enhancer–gene assignments based on correlation of H3K27ac and mRNA profiles across a wide array of tissues (Extended Data Fig. 9a). Top, an example locus encompassing *Tbx3*, *Tbx5*, and *Lhx5*, up to 25 enhancers are assigned to each of these three genes (blue, pink and brown boxes, Extended Data Fig. 9c). Genes showing fewer than five assigned enhancers are shown in grey. Bottom, heat maps showing meta-profiles of each gene's expression profile across tissues (red shades), along with the cumulative activity profile of its assigned enhancers (blue shades). **b**, Distribution of the number of enhancers assigned to developmental transcription factors (TFs) with biased expression in limb ( $P = 5 \times 10^{-19}$  versus housekeeping), forebrain ( $P = 8 \times 10^{-15}$ ), and heart ( $P = 3 \times 10^{-25}$ ) (two-sided Mann–Whitney tests). Box plots show median, interquartile values, range, and outliers (individual points). **c**, Complete spectrum of genes with at least one assigned enhancer, sorted by decreasing enhancer number. Limb-biased transcription factors are highlighted in green. **d**, Total number of enhancers (in all tissues analysed) assigned to each transcription factor in **c**, with the number of assigned enhancers predicted specifically in limb at E11.5 (dark green) or any other stage analysed (light green).

during development, taking advantage of chromatin signatures of distal enhancers and gene transcription measured across multiple tissues and time points of mouse development (Fig. 4 and Extended Data Figs 9, 10). We analysed correlations between H3K27ac chromatin immunoprecipitation followed by sequencing (ChIP-seq) and RNA sequencing (RNA-seq) datasets from twelve different mouse tissues at two or three embryonic or perinatal time points per tissue (<https://www.encodeproject.org/>) to assign each enhancer to its most likely target gene within the same topologically associated domain (TAD)<sup>26</sup> (Fig. 4, Extended Data Fig. 9a–c and Methods). We then used this framework to examine the average number of enhancers associated with genes expressed in three developmental tissues (limb, heart, and forebrain). Genes with limb-biased expression showed a median of three associated distal enhancers, versus a median of zero for housekeeping genes (Extended Data Fig. 9d, e). For the specific class of limb-biased genes encoding transcription factors, we observed an even more complex enhancer landscape, with a median of eight distinct enhancers per gene (Fig. 4b). Notably, some of these transcription factor genes were associated with more than ten tissue-specific limb enhancers with highly overlapping activity patterns in the same tissue (Fig. 4c, d and Methods). We observed similarly large numbers of potentially redundant enhancers near brain- and heart-specific transcription factor genes (Extended Data Fig. 10a, b). Even under stringent correlation thresholds, our analysis uncovered 1,058 genes associated with five or more enhancers showing putatively redundant activity patterns—that is, enhancers that are active in the same tissue (Extended Data Fig. 10c–f). These results indicate that developmentally expressed genes are commonly associated with multiple enhancers that show overlapping activity patterns, supporting the widespread existence of functionally redundant enhancers in mammalian genomes.

Studies of individual loci have identified examples of mammalian enhancers near the same gene with remarkably similar spatiotemporal activity patterns or functions<sup>15,27–32</sup>, reminiscent of invertebrate ‘shadow enhancers’<sup>8,9,33–35</sup>. The lack of marked morphological phenotypes in our enhancer deletion mouse models suggests that panels of mammalian enhancers with large degrees of redundancy act as a regulatory buffer for key developmental processes, thereby reducing the likelihood of severe consequences resulting from genetic or environmental challenges<sup>8</sup>. Although individual examples of enhancers whose loss leads to severe phenotypes have been described<sup>4,36</sup>, our findings suggest that redundancy is far more common. As indicated by the phenotypes observed in sensitized genetic backgrounds, our results suggest that pairs of enhancers act redundantly in organismal patterning, but additively in establishing gene expression levels. This observation is consistent with high-throughput loss-of-function screens in cultured cells, in which the disruption of individual enhancers leads to measurable gene expression changes but rarely results in the complete loss of target gene expression<sup>37</sup>. It appears plausible to assume that limited but specific contributions to overall gene expression levels are relevant for organismal fitness under specific pressures, thus subjecting enhancers to purifying selection over evolutionary time. Alternatively, additional tissue-specific functions may also explain the evolutionary constraints on these loci.

Our observations have implications for the interpretation of noncoding regulatory variants in relation to human phenotypes. Our findings suggest that many loss-of-function enhancer mutations will cause, at most, subtle phenotypes in humans. Thus, for many genetic loci, enhancer-associated disease phenotypes may be more likely to result from gain-of-function mutations that either expand enhancer activity<sup>38</sup> or alter the positions of enhancers relative to genes<sup>39</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 March; accepted 18 December 2017.

Published online 31 January 2018.

1. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
3. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**, 3646–3658 (2017).
4. Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
5. Menke, D. B., Guenther, C. & Kingsley, D. M. Dual hindlimb control elements in the *Tbx4* gene and region-specific control of bone size in vertebrate limbs. *Development* **135**, 2543–2553 (2008).
6. Shim, S., Kwan, K. Y., Li, M., Lefebvre, V. & Sestan, N. Cis-regulatory control of corticospinal system development and evolution. *Nature* **486**, 74–79 (2012).
7. Hay, D. *et al.* Genetic dissection of the  $\alpha$ -globin super-enhancer *in vivo*. *Nat. Genet.* **48**, 895–903 (2016).
8. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).
9. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
10. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
11. Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245–258 (2017).
12. Zeller, R., López-Ríos, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat. Rev. Genet.* **10**, 845–858 (2009).
13. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
14. Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
15. Attanasio, C. *et al.* Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013).
16. Osterwalder, M. *et al.* HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Dev. Cell* **31**, 345–357 (2014).
17. Rosin, J. M., Abassah-Oppong, S. & Cobb, J. Comparative transgenic analysis of enhancers from the human *SHOX* and mouse *Shox2* genomic regions. *Hum. Mol. Genet.* **22**, 3063–3076 (2013).
18. Cobb, J., Dierich, A., Huss-Garcia, Y. & Duboule, D. A mouse model for human short-stature syndromes identifies *Shox2* as an upstream regulator of *Runx2* during long-bone development. *Proc. Natl Acad. Sci. USA* **103**, 4511–4515 (2006).
19. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrughe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16**, 2813–2828 (2002).
20. Kawakami, Y. *et al.* Sall genes regulate region-specific morphogenesis in the mouse limb by modulating Hox activities. *Development* **136**, 585–594 (2009).
21. Min, H. *et al.* Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to *Drosophila* branchless. *Genes Dev.* **12**, 3156–3161 (1998).
22. Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
23. Hui, C. C. & Joyner, A. L. A mouse model of Greig cephalopolysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the *Gli3* gene. *Nat. Genet.* **3**, 241–246 (1993).
24. Lopez-Rios, J. *et al.* Gli3 constrains digit number by controlling both progenitor proliferation and BMP-dependent exit to chondrogenesis. *Dev. Cell* **22**, 837–848 (2012).
25. Ye, W. *et al.* A unique stylopod patterning mechanism by *Shox2*-controlled osteogenesis. *Development* **143**, 2548–2560 (2016).
26. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
27. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
28. Visel, A. *et al.* A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).
29. Lam, D. D. *et al.* Partially redundant enhancers cooperatively maintain mammalian *pomc* expression above a critical functional threshold. *PLoS Genet.* **11**, e1004935 (2015).
30. Yao, Y. *et al.* Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat. Genet.* **48**, 575–580 (2016).
31. Antosova, B. *et al.* The gene regulatory network of lens induction is wired through Meis-dependent shadow enhancers of *Pax6*. *PLoS Genet.* **12**, e1006441 (2016).
32. Will, A. J. *et al.* Composition and dosage of a multipartite enhancer cluster control developmental expression of *lh* (Indian hedgehog). *Nat. Genet.* **49**, 1539–1545 (2017).
33. Hong, J. W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
34. Barolo, S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays* **34**, 135–141 (2012).
35. Cannavò, E. *et al.* Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38–51 (2016).
36. Yanagisawa, H., Clouthier, D. E., Richardson, J. A., Charité, J. & Olson, E. N. Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. *Development* **130**, 1069–1078 (2003).
37. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
38. Lettice, L. A., Hill, A. E., Devenney, P. S. & Hill, R. E. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985 (2008).
39. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by National Institutes of Health grants R01HG003988, U54HG006997, R24HL123879 and UM1HL098166 (to A.V. and L.A.P.) and the University of Basel and the Novartis Foundation for Biomedical Research (to J.L.-R.). M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. We thank B. Ren for providing access to the ChIP-seq and RNA-seq data from ENCODE; J. Doudna for providing a plasmid containing a human-optimized Cas9 gene; W. Ye and Y. Chen for sharing the image of a *Shox2*-deficient limb skeleton (Fig. 3b); and the members of the L.A.P., A.V. and D.E.D. groups for technical advice and comments on the manuscript, in particular C. Spurrell and E. Kvon. Research was conducted at the E. O. Lawrence Berkeley National Laboratory and performed under Department of Energy Contract DE-AC02-05CH11231, University of California.

**Author Contributions** M.O., D.E.D., A.V., and L.A.P. conceived the study. M.O., D.E.D., B.J.M., S.Y.A., E.A.L., Y.Z., I.P.-F., C.S.P., M.K., T.H.G., Q.T.P., A.N.H., J.A.A., and V.A. performed the genome editing and mouse phenotyping studies. I.B. and M.O. devised the computational framework, and I.B. performed the correlative analysis. V.T. performed *in situ* hybridization under the supervision of J.L.-R. Y.F.-Y. conducted the ChIP-seq and RNA-seq data analysis. M.O., D.E.D., A.V., and L.A.P. wrote the manuscript with input from the remaining authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.E.D. ([dedickel@lbl.gov](mailto:dedickel@lbl.gov)), A.V. ([avisel@lbl.gov](mailto:avisel@lbl.gov)), or L.A.P. ([lapennacchio@lbl.gov](mailto:lapennacchio@lbl.gov)).

## METHODS

**Experimental design.** All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory (LBNL) Animal Welfare Committee. All mice used in this study were housed at the Animal Care Facility (ACF) at LBNL. Mice were monitored daily for food and water intake, and animals were inspected weekly by the Chair of the Animal Welfare and Research Committee and the head of the animal facility in consultation with the veterinary staff. The LBNL ACF is accredited by the American Association for the Accreditation of Laboratory Animal Care (AAALAC). Transgenic mouse assays and enhancer knockouts were performed in *Mus musculus* FVB strain mice. The following developmental stages were used in this study: embryonic day E10.5, E11.5, E12.5 and E18.5 mice. Animals of both sexes were used in the analysis. Sample size selection and randomization strategies were conducted as described below.

**Transgenic mouse assay selection and randomization.** Sample sizes were selected empirically on the basis of our previous experience of performing transgenic mouse assays for more than 2,000 total putative enhancers (VISTA Enhancer Browser: <https://enhancer.lbl.gov/>). Mouse embryos were excluded from further analysis if they did not contain the reporter transgene or if the developmental stage was not correct. All transgenic mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed. **Enhancer knockout selection and randomization.** Sample sizes were selected empirically on the basis of our previous studies<sup>15</sup>. All phenotypic characterization of knockout mice used a matched littermate selection strategy. All phenotyped mice described in the paper resulted from crossing heterozygous enhancer deletion mice together to allow the comparison of matched littermates of different genotypes. Embryonic samples used for *in situ* hybridizations, RNA-seq, and skeletal preparations were dissected blinded to genotype.

**In vivo transgenic reporter assays.** Enhancer names in this study are the unique identifiers used in the VISTA Enhancer Browser (<https://enhancer.lbl.gov/>; mm: originally identified in mouse; hs: originally identified in human). Transgenic results for most enhancers have been reported previously<sup>13–16</sup>. Newly tested enhancers (hs1586 at E10.5 and hs1262) were amplified from human genomic DNA and cloned into an hsp68-*lacZ* expression vector as previously described<sup>14</sup>. Genomic coordinates of all enhancers are listed in Supplementary Table 1. *LacZ* transgenic mouse assays were conducted as previously described<sup>14,40</sup>. To directly compare the activity domains between apparently redundant enhancers, enhancers were cloned, using Gateway (Thermo Fisher Scientific) or Gibson<sup>41</sup> methods, into an hsp68-based reporter vector similar to that described above, with the exception of a fluorescent reporter replacing *LacZ*. The enhancer–reporter combinations were generated as follows: mm1179-sfGFP, hs1586-mCherry, hs741-sfGFP and hs1262-mCherry. sfGFP is a fusion of Sun1 and 2×sfGFP as described<sup>42</sup> and localizes to the nuclear membrane. Mice carrying the individual fluorescent reporter transgenes were then generated via pronuclear injection (using FVB strain zygotes), and stable lines were established from founders showing reproducible reporter activity in the embryonic limb.

**Generation of enhancer knockout mice using CRISPR–Cas9.** Mouse strains lacking limb enhancer(s) or harbouring gene loss-of-function alleles were generated using *in vivo* CRISPR–Cas9 editing, as previously described, with only minor modifications<sup>43,44</sup>. Pairs of single guide RNAs (sgRNAs) targeting genomic sequence 5' and 3' to the sequence to be deleted were designed using CHOPCHOP<sup>45</sup> (see Supplementary Table 1 for sgRNA sequences and coordinates of deleted regions). Knockout mice were engineered as described previously<sup>46</sup> using a mix containing Cas9 mRNA (final concentration of 100 ng/μl) and two sgRNAs (25 ng/μl each) in injection buffer (10 mM Tris, pH 7.5; 0.1 mM EDTA). This mix was injected into the cytoplasm of single-cell FVB strain mouse embryos. Founder (F0) mice were genotyped using PCR with High Fidelity Platinum Taq Polymerase (Thermo Fisher Scientific) to identify those with the desired non-homologous end joining (NHEJ)-generated deletion breakpoints (see Extended Data Figs 1a–j, 5a–c, 7a, d and Supplementary Table 3 for genotyping strategy, primer sequences and PCR amplicons). Sanger sequencing was used to identify and confirm deletion breakpoints in F0 and F1 mice (Extended Data Figs 1a–j, 5a–c, 7a, d). Unless noted otherwise, mice homozygous-null for the targeted limb enhancers showed normal pre- and postnatal viability and appeared outwardly normal. For iterative CRISPR–Cas9 genome editing, fertilized mouse eggs harbouring the primary deletion were collected and injected with sgRNAs targeting the secondary enhancer for deletion. Only those founder lines harbouring both deletions on the same haplotype were analysed further.

**In situ hybridization and skeletal preparations.** To assess spatial changes in gene expression in mouse embryonic limbs, whole mount *in situ* hybridization using digoxigenin-labelled antisense riboprobes was carried out as previously described<sup>46</sup>. Forelimbs and hindlimbs from at least three independent embryos were analysed for each genotype (including wild-type littermate controls). Mouse embryonic skeletons at E18.5 were stained with Alcian blue and Alizarin red to differentiate

cartilage (blue) and bone (red) using standard methods<sup>47</sup>. For comparison of limb skeletons from enhancer knockout embryos and wild-type littermates, general parameters such as bone number, shape, length, position or mineralization were assessed. Embryonic limbs and limb skeletons were imaged, and skeletal elements were measured, using a Leica MZ16 stereo-microscope coupled to a Leica DFC300F or DFC420 digital camera. Brightness and contrast were adjusted uniformly using Photoshop CS5. Measurements of the ossified portions of humerus and femur (stylopodial elements) were normalized to those of the ulna and tibia (related zeugopodial elements), respectively (as shown in Figs 2d, 3b and Extended Data Figs 5i, 8c).

**Quantitative real-time PCR (qPCR) and RNA-seq.** RNA was isolated from microdissected forelimbs or hindlimbs of mouse embryos at E11.5 using the Ambion RNeasy Total RNA Isolation Kit (Life Technologies) according to the manufacturer's instructions. For qPCR, RNA was treated with RNase-free DNase (Promega) and reverse transcribed using SuperScript III (Life Technologies) with random hexamer or poly-dT priming according to the manufacturer's instructions. qPCR was performed on a LightCycler 480 (Roche) using KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) according to the manufacturer's instructions. qPCR primers (listed in Supplementary Table 4) were designed *in silico* using Primer3 (<http://primer3.wi.mit.edu/>), and amplicons span exon–exon junctions in order to prevent amplification of genomic DNA. Relative gene expression levels were calculated using the  $2^{-\Delta\Delta C_t}$  method<sup>48</sup>, normalized to the *Actb* housekeeping gene, and the mean of wild-type control samples was set to 1.

For RNA-seq, RNA samples were treated with DNase (TURBO DNA-free Kit, Life Technologies), and RNA quality was verified using a 2100 Bioanalyzer (Agilent) with an RNA 6000 Nano Kit (Agilent). RNA-seq libraries were generated using the TruSeq Stranded mRNA Sample Prep Kit (Illumina), following the manufacturer's instructions, and purified, eluted, and quantified as described previously<sup>49</sup>. RNA-seq libraries were pooled (four per lane) and sequenced using single end 50-bp reads on a HiSeq 4000 (Illumina).

**Immunofluorescence.** Mouse embryonic limbs at E10.5, E11.5 or E12.5 were dissected in cold PBS and fixed in 4% PFA for 2–3 h. Following incubation in a sucrose gradient and embedding in a 1:1 mixture of 30% sucrose and optimum cutting temperature solution, sagittal 10-μm frozen sections were cut using a cryostat. Cryosections were incubated overnight with the following primary antibodies: chicken anti-GFP (1:500, Thermo Fisher Scientific, A10262), rabbit anti-mCherry (1:1,000, Thermo Fisher Scientific, PA5-34974) and goat anti-Sox9 (1:500, R&D Systems, AF3075). Goat-anti chicken, goat anti-rabbit and donkey anti-goat secondary antibodies conjugated to Alexa Fluor 488, 568, 594 or 647 (1:1,000, Thermo Fisher Scientific) were used for detection. Hoechst 33258 (Sigma-Aldrich) was used to counterstain nuclei. Fluorescent images were acquired using a Zeiss AxioImager fluorescence microscope in combination with a Hamamatsu Orca-03 camera. Brightness and contrast were adjusted uniformly using Photoshop CS5.

**ChIP-seq.** For each of six single enhancer knockout lines, ChIP-seq to H3K27ac was performed using a protocol optimized for mouse embryonic tissues<sup>50</sup>. In brief, forelimb buds from ten wild-type embryos (four biological replicates) and ten enhancer knockout embryos (at least two biological replicates) were dissected at E11.5, formaldehyde crosslinked, and sheared using a Diagenode Bioruptor Sonicator. After pre-clearing, chromatin was incubated with anti-H3K27ac antibody (Active Motif cat no. 39133) for 2 h at 4 °C. Freshly rinsed Dynabeads (1:1 protein A:protein G mix) were then added to the antibody-treated chromatin, and immunoprecipitation was performed on a rotator for 30 min at 4 °C. Libraries were prepared using the Illumina TruSeq DNA sample prep kit following the manufacturer's instructions with minor modifications. Library quality was assessed using a 2100 Bioanalyzer with the High Sensitivity DNA Kit (Agilent), and quantification was performed using a Qubit Fluorometer with the dsDNA HS Assay Kit (Life Technologies). ChIP-seq and input libraries were pooled and sequenced via single-end 50-bp reads on a HiSeq 2000 or 4000 (Illumina).

**RNA-seq and ChIP-seq analysis.** Analysis of ChIP-seq and RNA-seq data from limb enhancer knockout and related wild-type control samples was performed as follows: CASAVA v1.8.0 (Illumina) was used to demultiplex data, and reads with CASAVA 'Y' flag (purity filtering) were discarded. For each sample, between 12 million and 55 million (ChIP-seq) or 23 million and 71 million (RNA-seq) reads were obtained following quality filtering and adaptor trimming using cutadapt\_v1.1 (<https://cutadapt.readthedocs.io/>) with parameter '-m 25 -q 25'. Mouse genome sequence (mm9) and gene annotations were retrieved from the iGenomes repository ([https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html)).

To align the RNA-seq reads to the mouse reference genome and transcriptome, we used Tophat v2.0.6<sup>51</sup>, and the reads mapping to UCSC known genes were counted by HTSeq<sup>52</sup>. Genes with counts per million (CPM) > 1 in at least two samples were processed for further differential gene expression analysis comparing enhancer knockout and wild-type control samples using edgeR<sup>53</sup>. In each case, the

top 100 differentially expressed genes, sorted by false discovery rate (FDR), are listed in Supplementary Tables 5–7.

For read mapping and peak calling of ChIP-seq datasets, bowtie<sup>54</sup> (version 0.12.8) with parameter ‘-m 1 -v 2’ and MACS<sup>55</sup> (version 1.4.2) with parameter ‘-mfold = 10,30 -nomodel -p 0.0001’ were used, respectively. Biological replicates were combined using MSPC<sup>56</sup>, with the following parameters: -r biological -s 1E-10 -W 1E-6 -m Highest -c 2. The predicted enhancer intervals were assigned the best *P* value (as defined by MACS<sup>55</sup>) among the overlapping peaks.

**ENCODE ChIP-seq data analysis.** Raw data were downloaded from the Data Coordination Center of the ENCODE project (<https://www.encodeproject.org/>; see Supplementary Table 8 for the complete list of sample identifiers). Short reads were aligned to the mm10 assembly of the mouse genome using bowtie<sup>54</sup>, with the following parameters: -a -m 1 -n 2 -l 32 -e 3001. Peak calling was performed using MACS v1.4<sup>55</sup>, with the following arguments: -gsize = mm-bw = 300 -nomodel -shiftsize = 100. Experiment-matched input DNA was used as a control.

**ENCODE RNA-seq data analysis.** Raw data were downloaded from the ENCODE Data Coordination Center (<https://www.encodeproject.org/>; see Supplementary Table 8 for the complete list of sample identifiers). Short reads were aligned to the mm10 assembly of the mouse genome using Tophat v2.0.8<sup>57</sup> and Gencode vM3<sup>58</sup> as the reference transcriptome. Cuffnorm v2.2.1<sup>51</sup> was run to quantify transcripts across conditions using the Gencode vM3<sup>58</sup> transcriptome as the reference and setting -library-norm-method to geometric. Only genes with a level of expression of at least one RPKM (reads per kilobase of exons per million mapped reads) in at least one of the considered conditions were included in further analyses. Small and non-coding RNAs were excluded by retaining only those genes with a Gencode biotype<sup>58</sup> supporting protein-coding functionality.

**Classifying genes by tissue-biased patterns of expression.** For each protein-coding gene in the mouse genome, the expression variability across the twenty-nine ENCODE RNA-seq experiments from multiple tissues and developmental time points was evaluated using two metrics: a measure of tissue-specificity ( $\tau$ )<sup>59</sup> ranging from 0 (consistent expression across all conditions) to 1 (expression in one single condition); and a measure of relative expression in a condition of interest (for example, limb at E11.5). Given a gene, the latter was defined as the difference between the percentile of expression of the gene in the given condition and the median percentile of expression across all the samples. A large positive number indicates a gene that is much more expressed in the condition of interest than the average.

Tissue-biased genes were defined as showing  $\tau \geq 0.7$  and relative expression higher than the 95th percentile. Housekeeping genes were defined as having  $\tau \leq 0.4$  and relative expression between the 5th and 95th percentiles. The complete lists of genes assigned to each category are available in Supplementary Table 9.

**Gene classification based on pre-specified functional categories.** Tissue-biased developmental transcription factors (sometimes referred to as tissue-specific transcription factors) were defined as genes with biased expression in a given tissue (see previous section), associated with abnormal developmental phenotypes in the same tissue (terms extracted from the Mouse Genome Informatics (MGI) database<sup>60</sup>, listed in Supplementary Table 10) and annotated as a transcription factor under the terms GO:0003700 or GO:0003705 in the Gene Ontology (GO)<sup>61</sup>. Annotations were downloaded from GO and MGI on July 7, 2016.

**Topologically associated domains.** TAD coordinates<sup>26</sup> estimated from mouse embryonic stem cell Hi-C data were downloaded from <http://chromosome.sdsc.edu/mouse/hi-c/download.html>. Coordinates were converted from mm9 to mm10 using liftOver<sup>62</sup>.

**A statistical framework defining enhancer-promoter associations genome-wide.** A list of putative enhancer regions was first defined as follows: after excluding any region annotated to the mitochondrial or any random chromosome, the BED coordinates of the H3K27ac peaks across the twenty-nine conditions (different combinations of tissue and developmental stage as defined by the ENCODE consortium, see ‘ENCODE ChIP-seq data analysis’ above) were merged using the mergeBed utility from BEDTools v2.17.0<sup>63</sup>. For a more robust signal estimation (see below), regions shorter than 500 bp were enlarged to 1 kb from their central coordinate. Promoters, defined as regions within 2.5 kb of the transcriptional start sites of genes annotated in Gencode vM3<sup>58</sup>, were then excluded using subtractBed from BEDTools v2.17.0<sup>63</sup>. After that, any remaining region shorter than 1 kb was excluded. Uniquely aligned, de-duplicated reads were then used to quantify the H3K27ac signals at each region, for each one of the 29 conditions. These signals were measured using the coverageBed utility from BEDTools v2.17.0<sup>63</sup>, normalized to RPKM (according to the sequencing depth of each specific sample), and log<sub>2</sub>-transformed. The resulting list of 74,366 predicted enhancers and their corresponding H3K27ac signal quantifications, along with the mRNA expression measurements for the protein-coding genes (as defined in ‘Classifying genes by tissue-biased patterns of expression’), were used as input for the statistical

framework described below. The main steps of the approach are also outlined in Extended Data Fig. 9b.

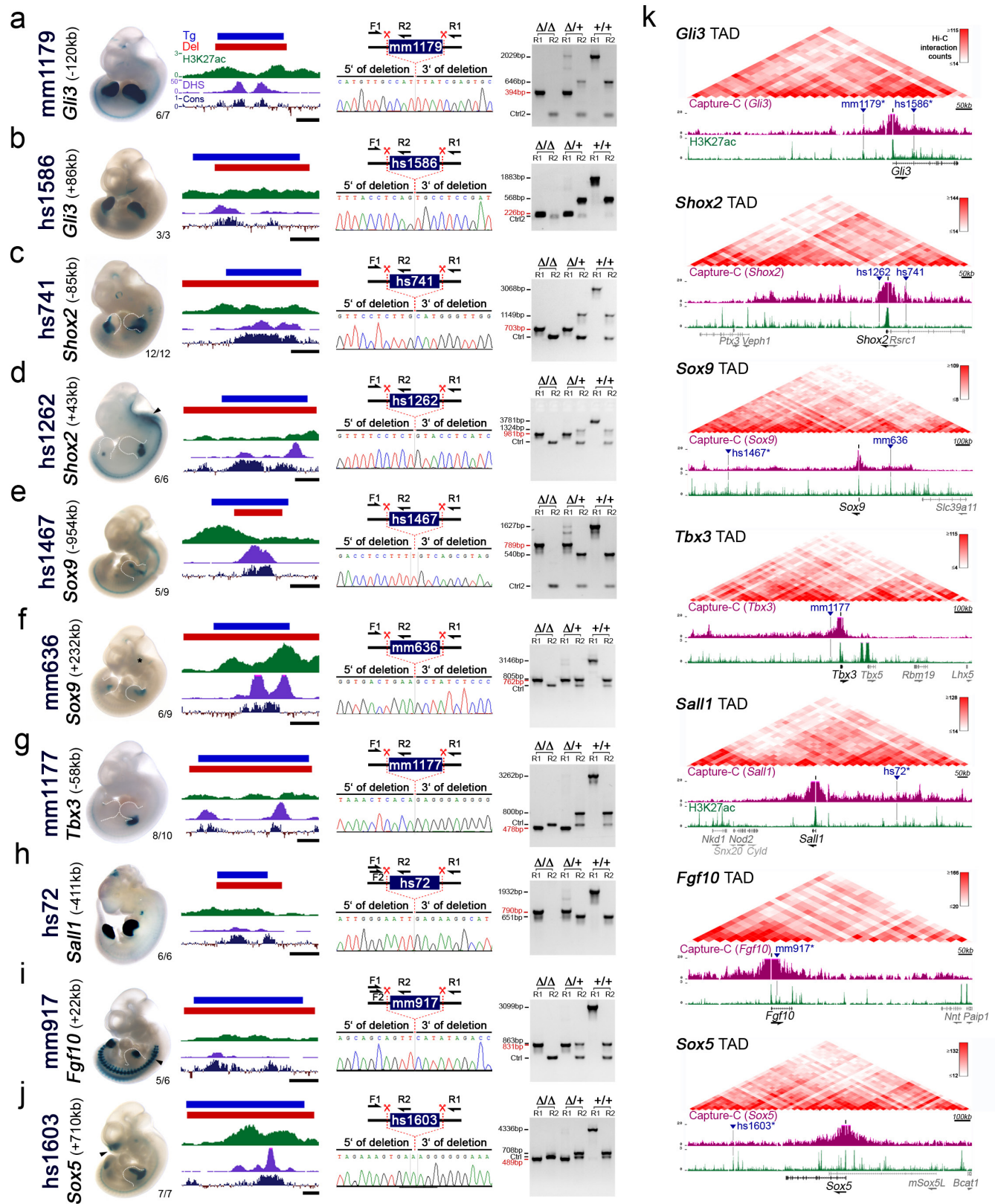
For each previously defined TAD in the mouse genome<sup>26</sup>, we retrieved all of the enhancers predicted and the genes expressed in at least one of the twenty-nine conditions considered that fell within that TAD. Pairwise correlations between all possible enhancer-gene combinations within the TAD were then evaluated by calculating Spearman's rank correlation coefficient (SCC) between the H3K27ac pattern of enrichment at the enhancer and the mRNA expression of the gene across the conditions. Each putative enhancer was initially assigned to the gene showing the highest SCC value (in the very rare case of ties, all of the genes showing the same SCC value were assigned to the enhancer). After that, a null distribution of SCC values was estimated empirically, by pairing the enhancer with 1,000 randomly picked genes from the same chromosome. The *z*-score for the correlation coefficient was then calculated by subtracting the mean and dividing by the standard deviation estimated from the empirical null. The corresponding *P* value was calculated using the pnorm function in R. Finally, only those putative enhancers showing a *P* value  $\leq 0.05$  and a SCC  $\geq 0.25$  were retained, resulting in a set of 34,882 enhancers with an assigned target (Supplementary Table 11). Considering the entire, genome-wide set of pairwise associations, a *P* = 0.05 corresponds to a Benjamini-Hochberg corrected FDR of 0.087. This analysis resulted in the assignment of one or more putative enhancers to 9,365 protein-coding genes (Supplementary Table 12). To define a set of genes with many redundant enhancers, we considered enhancers as redundant only if they were associated with the same gene by correlation and showed a strong peak of H3K27ac in the same exact tissue under examination (for example, both enhancers are active in limb and linked to the *Gli3* gene). Although this correlative approach may result in a subset of false-positive assignments for individual genes, it enables an approximation of both regulatory complexity and potential enhancer redundancy across the entire genome. We found 1,276 genes that showed multiple assigned enhancers such that at least five of the enhancers were all active in the same tissue (limb, heart or brain). We then used a permutation scheme to directly evaluate the statistical robustness of this conclusion (that is, 1,276 genes with 5 or more redundant enhancers in either developing limbs, heart or forebrain), which considered increasingly higher correlation values between the activity of putative enhancers and expression of genes (Extended Data Fig. 10c–f). By re-shuffling the expression values of each gene across conditions (100 genome-wide permutations), we estimated the FDR of observing a gene with five or more enhancers attached to it, for increasingly larger correlation coefficients. Each permutation consisted of the same enhancers and genes, in which the H3K27ac values were left as in the actual data whereas the RNA expression values of the genes across the different samples were randomly reshuffled. For each genome-wide permuted matrix, the entire statistical approach described above was re-run and a map of enhancer-promoter associations was generated. For each value of Spearman's correlation coefficient (0.25 to 0.75, with a 0.01 step) the number of genes showing five or more enhancers in the permuted data was calculated. The average across the 100 iterations was then computed and used for FDR estimation. This was calculated as the average number of genes showing five or more enhancers across the permuted data, over the number of genes derived from the actual data.

**Statistical analysis.** Statistical analyses are described in detail in the Methods sections above. Whenever a *P* value is reported in the text, the statistical test is also indicated. Unless specified otherwise, all the statistics were estimated and plots drawn using the statistical computing environment R (<https://www.r-project.org>) or GraphPad Prism 7 software.

**Data availability.** ChIP-seq and RNA-seq datasets are available in the NCBI GEO database with the accession code GSE93730. Additional data supporting the findings of this study are available from the corresponding authors upon reasonable request.

40. Kothary, R. *et al.* Inducible expression of an *hsp68-lacZ* hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).
41. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
42. Mo, A. *et al.* Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
43. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379 (2013).
44. Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Protocols* **9**, 1956–1968 (2014).
45. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
46. Kvon, E. Z. *et al.* Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642 (2016).

47. Ovchinnikov, D. Alcian blue/alizarin red staining of cartilage and bone in mouse. *Cold Spring Harb. Protoc.* **2009**, prot5170 (2009).
48. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative  $C_T$  method. *Nat. Protocols* **3**, 1101–1108 (2008).
49. Dickel, D. E. *et al.* Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat. Commun.* **7**, 12923 (2016).
50. Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
51. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
52. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Jalili, V., Matteucci, M., Masseroli, M. & Morelli, M. J. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**, 2761–2769 (2015).
57. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
58. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
59. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
60. Bult, C. J., Eppig, J. T., Blake, J. A., Kadin, J. A. & Richardson, J. E. Mouse genome database 2016. *Nucleic Acids Res.* **44**, D840–D847 (2016).
61. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
62. Speir, M. L. *et al.* The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.* **44**, D717–D725 (2016).
63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
64. Gordon, C. T. *et al.* Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. *Hum. Mutat.* **35**, 1011–1020 (2014).
65. Li, Q. *et al.* A Gli silencer is required for robust repression of gremlin in the vertebrate limb bud. *Development* **141**, 1906–1914 (2014).

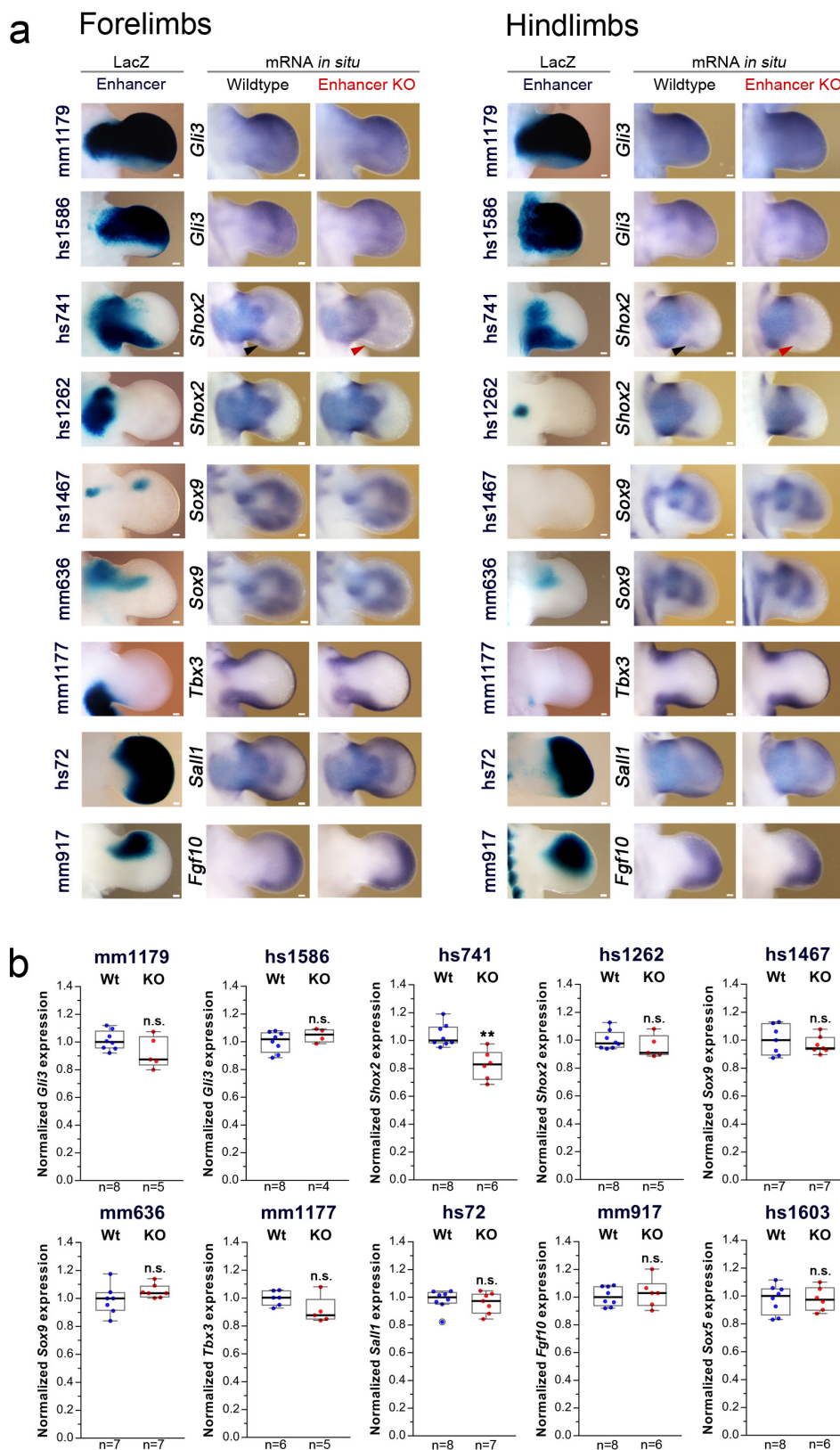


Extended Data Figure 1 | See next page for caption.

### Extended Data Figure 1 | CRISPR deletion of ten limb enhancers and regulatory interaction landscape of associated target genes.

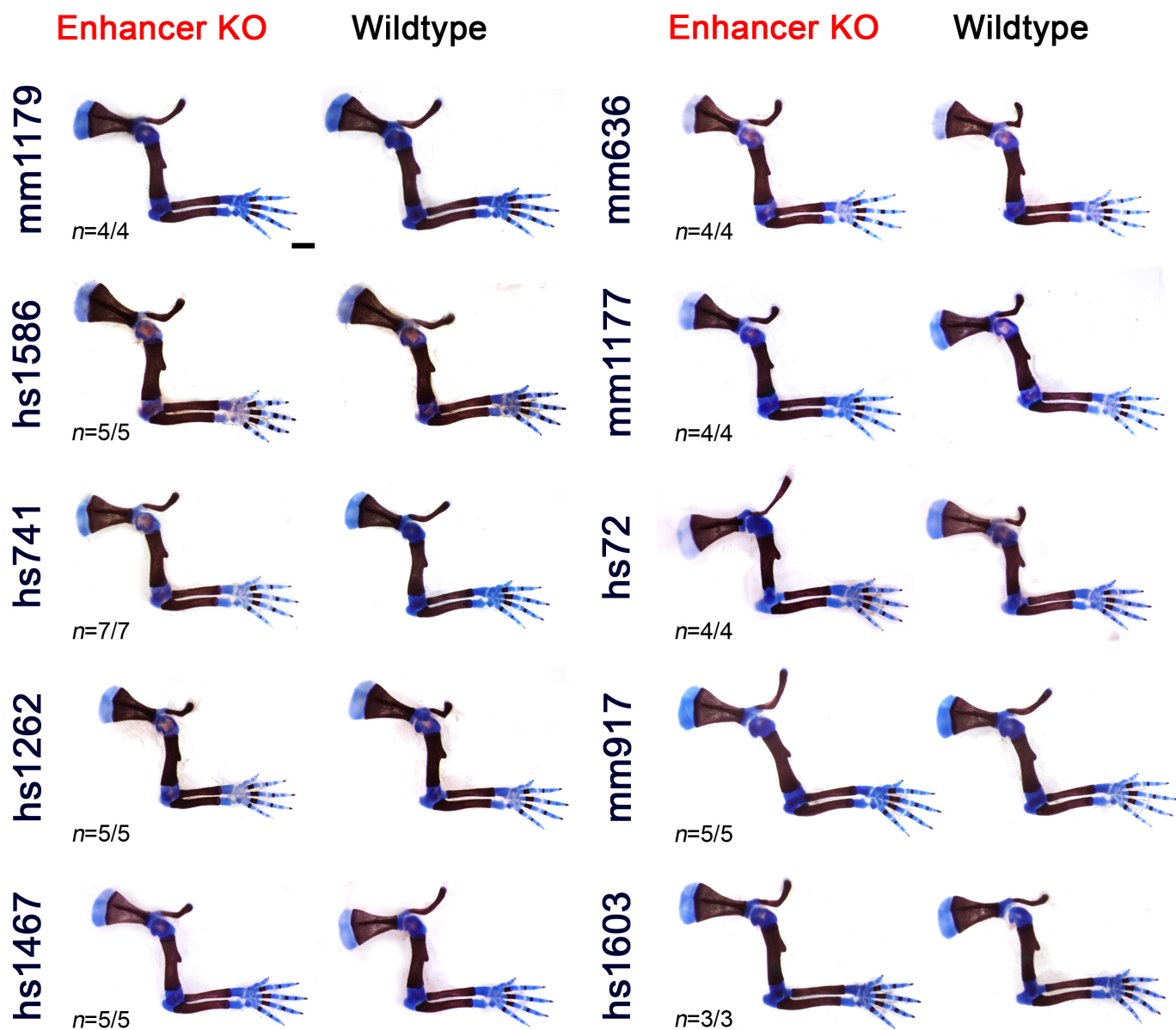
**a–j**, Left, representative activity patterns of the selected enhancers in mouse embryos at E11.5 (VISTA enhancer browser)<sup>13</sup> and the respective genomic enhancer regions tested in transgenic assays (Tg, blue bar), along with the regions deleted in enhancer knockout mice (Del, red bar). Corresponding H3K27 acetylation patterns (green) in wild-type mouse embryonic forelimbs at E11.5 (this study) are depicted with open chromatin (ENCODE DHS in forelimbs at E11.5, purple) and the Placental Mammal basewise conservation track by PhyloP (Cons, blue/red). Scale bars, 500 bp. VISTA enhancer IDs (mm and hs numbers) are indicated on the left, with the distance of the enhancer from the transcriptional start site of the predicted target gene in the mouse genome. Numbers at the bottom right of each embryo indicate the reproducibility of the enhancer reporter assay. Arrowheads mark additional activity domains (other than limb): hs1262 (hindbrain, reproducibility: 5/6, also shown previously<sup>17</sup>), mm917 (dorsal root ganglion, 7/7) and hs1603 (nose, 7/7; and branchial arch, 5/7). Asterisk indicates potential craniofacial enhancer activity for mm636, which was observed in 3 of 9 embryos<sup>64</sup>. Right, PCR validation strategy and results for enhancer knockout lines. Red scissors indicate

CRISPR-mediated deletion breakpoints. PCR was used to detect the wild-type (+) and enhancer deletion ( $\Delta$ ) alleles. Below, Sanger sequencing traces show the deletion breakpoints (indicated by the dashed line) for the enhancer knockout alleles. PCR genotyping results are shown with amplicon sizes indicated on the left (enhancer deletion allele in red). Primers (Ctrl or Ctrl2) amplifying an unrelated genomic region were included as a PCR positive control. See Supplementary Table 3 for all primer sequences and related PCR product sizes. **k**, Top, Hi-C interaction heat maps of topologically associated chromatin domains (mouse embryonic stem cell TADs)<sup>26</sup>. Bottom, selected enhancers (blue triangles) and their predicted target genes (TSS indicated as black bar). The Capture-C UCSC browser track (purple) illustrates three-dimensional chromatin interaction profiles from E11.5 embryonic limbs (3-kb window) using promoters of the predicted enhancer target genes as anchor points<sup>22</sup>. H3K27ac enrichment (green) in wild-type forelimbs at E11.5 (this study) is shown below. Six of the ten enhancers selected for deletion analysis display local Capture-C enrichment (\*), indicating physical interaction with the predicted target gene promoter at E10.5 or E11.5, based on the stringent statistical approach (95th percentile threshold) applied in the original study<sup>22</sup>. Other genes present in the TAD are shown in grey.



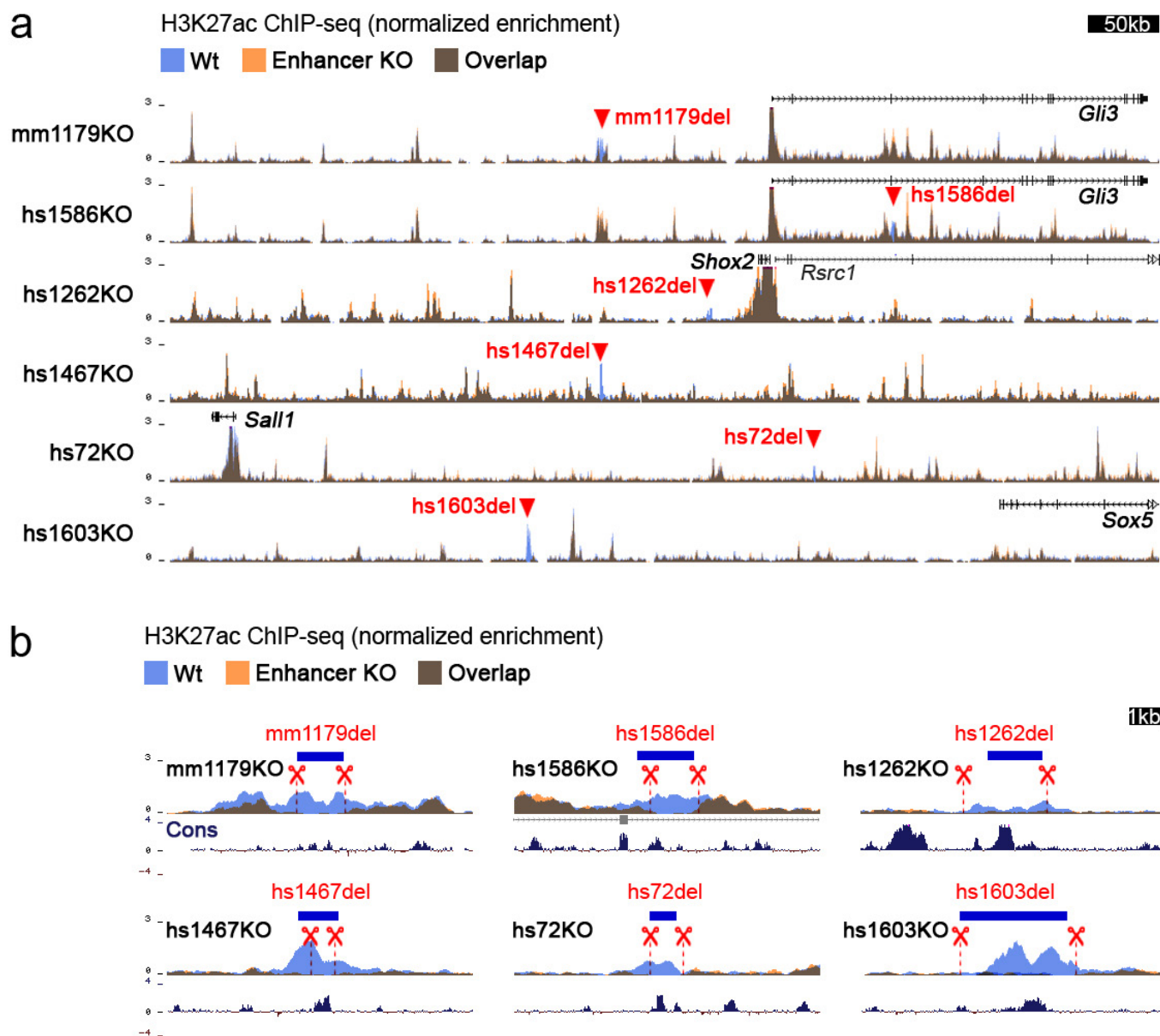
**Extended Data Figure 2 | No major differences in expression of predicted target genes in individual enhancer knockouts. a,** Spatial enhancer activity domains (LacZ, see also Fig. 1b) are compared to mRNA expression domains (by *in situ* hybridization) of the predicted target genes in embryonic forelimbs and hindlimbs at E11.5. No significant changes in expression patterns were observed in enhancer knockouts compared to wild-type limbs, except in limbs lacking hs741, where a small subdomain of target gene expression was lost (red arrowhead marks loss of the posterior *Shox2* domain in the distal limb, compared with black arrowhead

in wild type). Transcript distribution was reproduced in at least  $n = 3$  independent biological replicates. **b,** Quantitative real-time PCR using limbs of homozygous null (KO, red dots) and wild-type (Wt, blue dots) embryos at E11.5 reveals lack of significantly downregulated transcript levels of predicted enhancer target genes in nine out of ten cases. Box plots indicate median, interquartile values, range and individual biological replicates. Outliers are shown as circled data points.  $**P = 0.0012$ , unpaired, two-tailed *t*-test. n.s., not significant. Scale bars, 100  $\mu$ m.



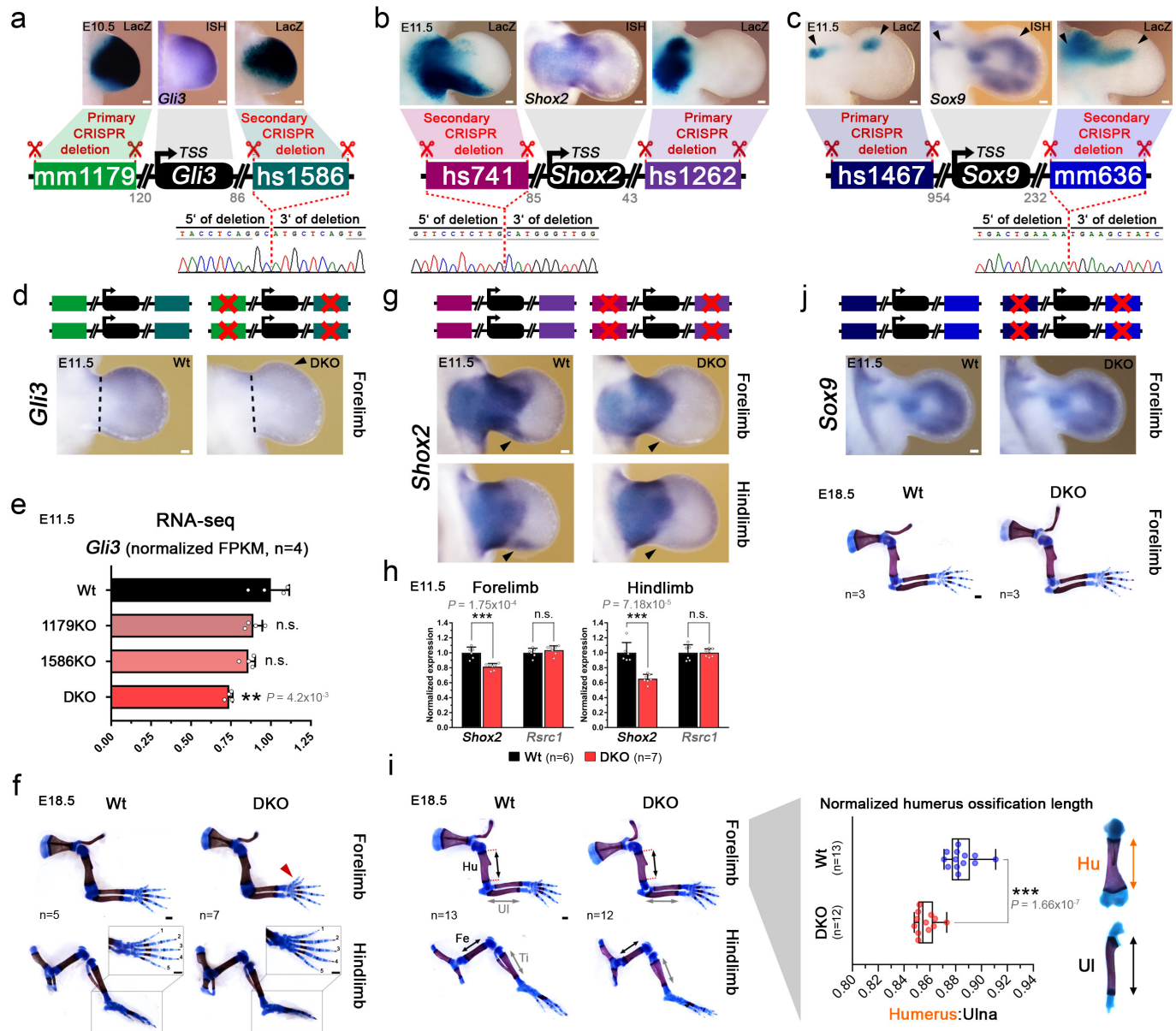
**Extended Data Figure 3 | Absence of obvious morphological abnormalities in limb enhancer knockouts.** Side-by-side comparison of enhancer knockout limb skeletons and wild-type littermate controls at E18.5. Neither forelimbs (this figure) nor hindlimbs (data not shown) of the enhancer knockout lines revealed any obvious morphological

differences in comparison to wild-type littermates. Cartilage is stained blue and bone dark red. The number of embryos with normal limb phenotypes over the total number of homozygous-null embryos examined is shown in the bottom left. *n* represents number of independent biological replicates with similar results. Scale bar, 1 mm.



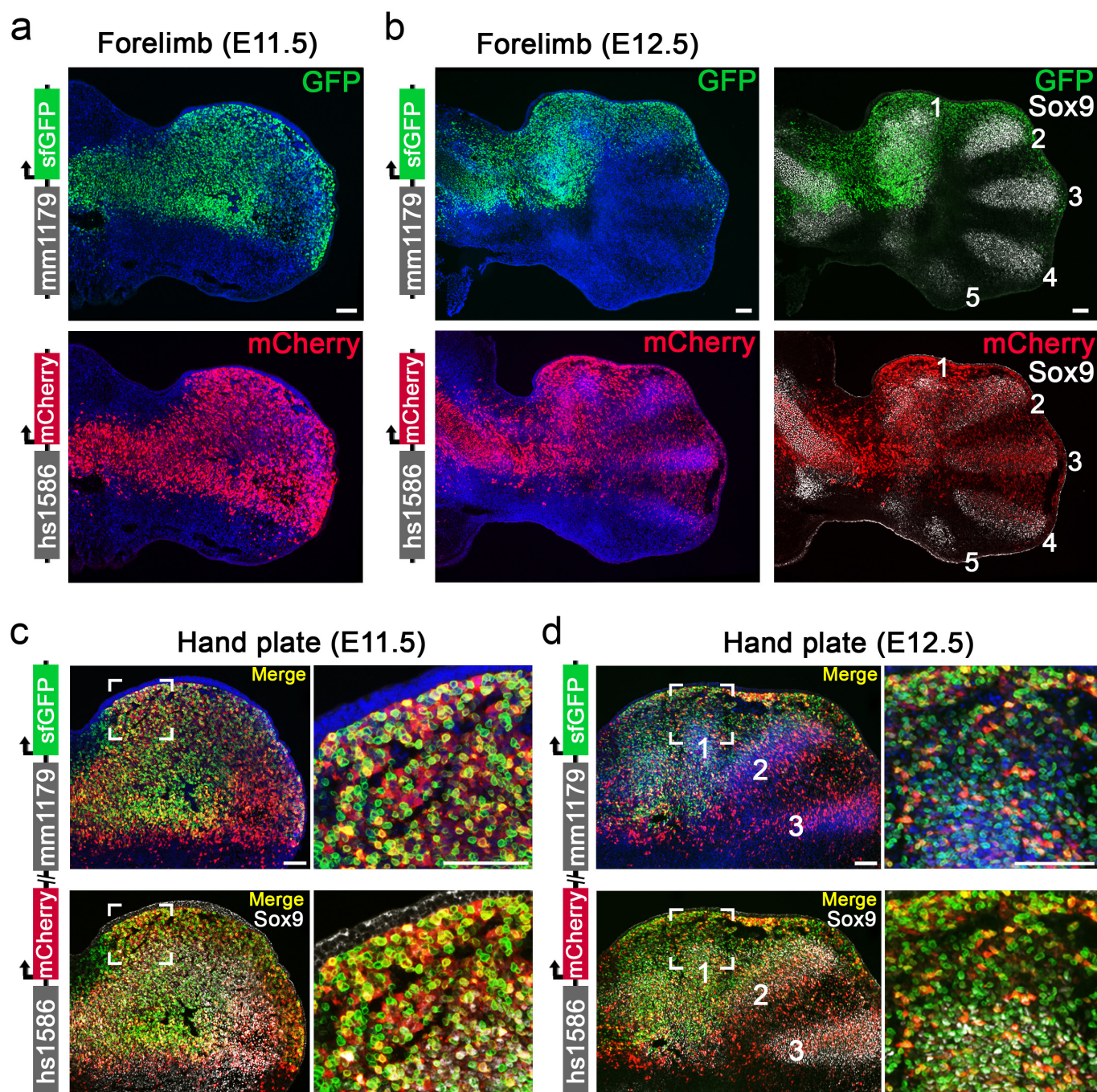
**Extended Data Figure 4 | Absence of compensatory enhancer signatures in limbs of enhancer knockout embryos. a**, Layered ChIP-seq H3K27 acetylation (ac) profiles surrounding the deleted enhancers and from wild-type (blue,  $n = 4$  independent biological replicates) and enhancer knockout embryos (orange, at least  $n = 2$  biological replicates). For all samples, E11.5 forelimb was profiled. For display, replicates were merged using bigWigMerge (UCSC tools) and normalized. Red triangles indicate

the positions of individual enhancer deletions. **b**, H3K27ac enrichments in targeted regions marked by red triangles in **a**, showing the absence of H3K27ac at the deletion site in individual enhancer knockout (orange) compared to wild-type (blue) samples. Blue bars indicate locations of enhancer sequences. Dashed red lines demarcate the regions deleted by CRISPR. Vertebrate basewise conservation track by PhyloP (Cons) is shown.



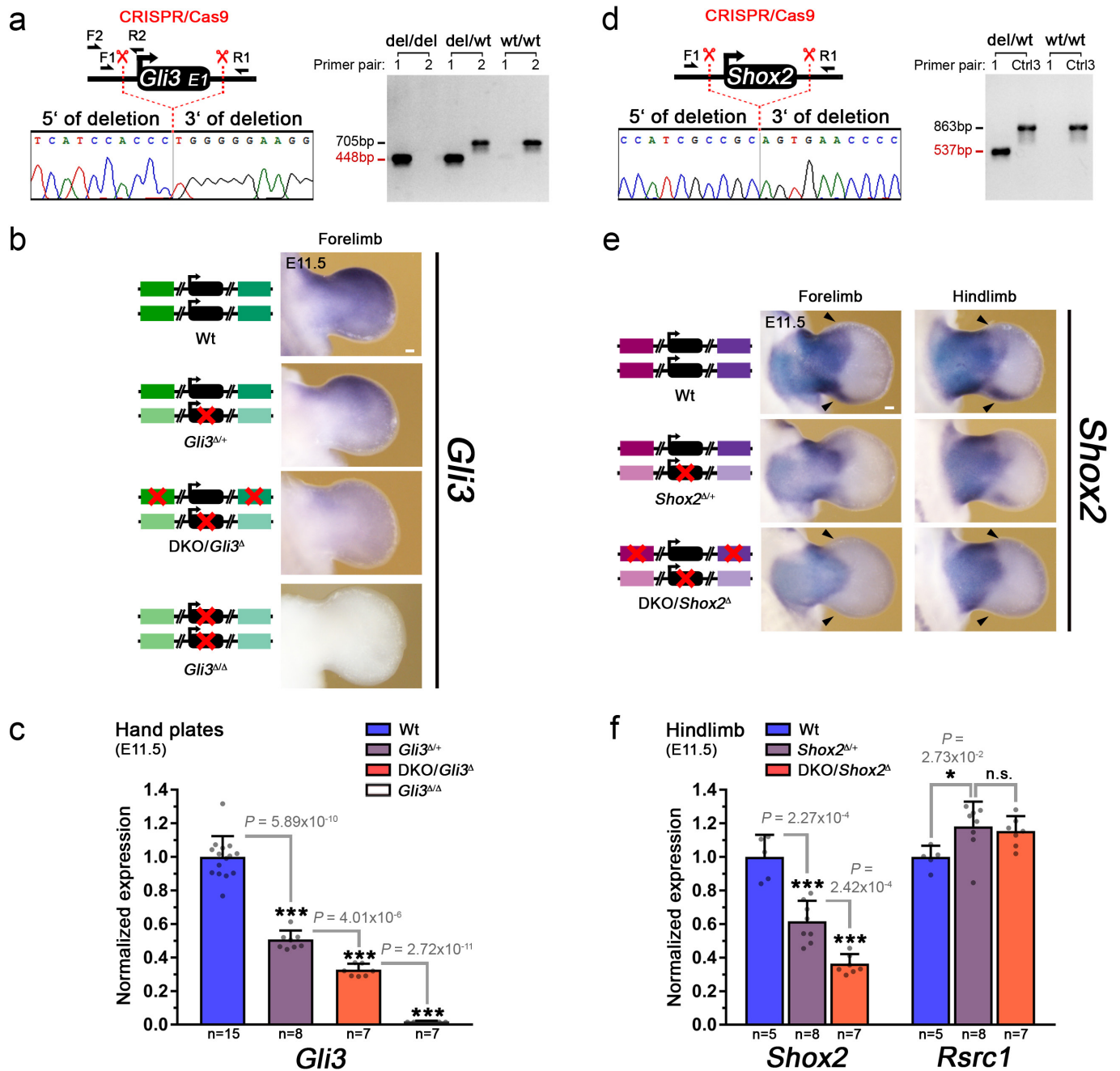
**Extended Data Figure 5 | Transcriptional and phenotypic impact of dual enhancer deletions engineered by iterative CRISPR–Cas9 genome editing.** **a–c**, Top, enhancer pairs with overlapping limb activities (LacZ), coinciding with domains of predicted target gene expression visualized by *in situ* hybridization (ISH). For *Sox9* enhancers, black arrowheads indicate overlapping domains. Schematics, double enhancer deletion strategy to delete the three enhancer pairs with overlapping activity (see Methods). Grey numbers indicate enhancer distance (kb) from the TSS. Bottom, Sanger sequencing verification of the secondary enhancer deletion. Deletion breakpoint is marked by the dashed line. Grey horizontal bars indicate bases present in the primary deletions (single enhancer knockout lines, see Extended Data Fig. 1a–j). **d**, *Gli3* transcript distribution *in situ* hybridization in wild-type (Wt) and mm1179/hs1586 DKO embryos. Arrowhead points to reduced *Gli3* transcript in the anterior limb mesenchyme. Dashed line indicates dissected hand plate for RNA-seq. **e**, RNA-seq confirmed significantly reduced *Gli3* expression in hand plates of DKO embryos but not individual enhancer knockout embryos (compared to wild-type hand plates). **f**, Unaffected hindlimb morphology in mm1179/hs1586 DKO embryos. Red arrowhead points to digit 1

duplication in forelimbs (see also Fig. 2). **g**, *Shox2* expression (*in situ* hybridization) in forelimbs and hindlimbs of hs741/hs1262 DKO embryos. The distal-posterior domain (arrowhead) is dependent on hs741 (Extended Data Fig. 2a). **h**, Reduced *Shox2* expression in forelimbs and hindlimbs of hs741/hs1262 DKO embryos (qPCR). Expression of the nearby *Rsrc1* gene was unchanged. **i**, Left, representative limb skeletons of wild-type and hs741/hs1262 DKO embryos. Hu, humerus; UL, ulna; Fe, femur; Ti, tibia. Right, mild but significant reduction in humerus ossification length (double arrows) in hs741/hs1262 DKO limb skeletons.  $***P = 1.66 \times 10^{-7}$  (two-tailed, unpaired *t*-test). **j**, Absence of evident differences in *Sox9* expression or skeletal abnormalities in embryos lacking both the hs1467 and mm636 enhancers near *Sox9*. For *in situ* hybridization, transcript distribution was reproduced in at least  $n = 3$  independent biological replicates. *n* represents number of independent biological replicates with similar results. For bar graphs and boxplots, individual biological replicates are shown as data points. Bar graphs illustrate mean and s.d. Box plot indicates median, interquartile values and range.  $***P < 0.001$ ;  $**P < 0.01$  (two-tailed, unpaired *t*-test). n.s., not significant. Scale bars, 100  $\mu$ m (white) and 500  $\mu$ m (black).



**Extended Data Figure 6 | Cellular resolution of redundant *Gli3* enhancer activities at the onset of digit formation.** **a, b**, Individual *Gli3* enhancer activities as detected by immunofluorescence (mm1179, green; hs1586, red) in forelimbs of transgenic reporter embryos. Sox9 (grey) marks chondrogenic progenitors of the mesenchymal condensations forming digit primordia (digits 1–5, from anterior to posterior). **c, d**, Co-localization of mm1179 and hs1586 enhancer activities in hand

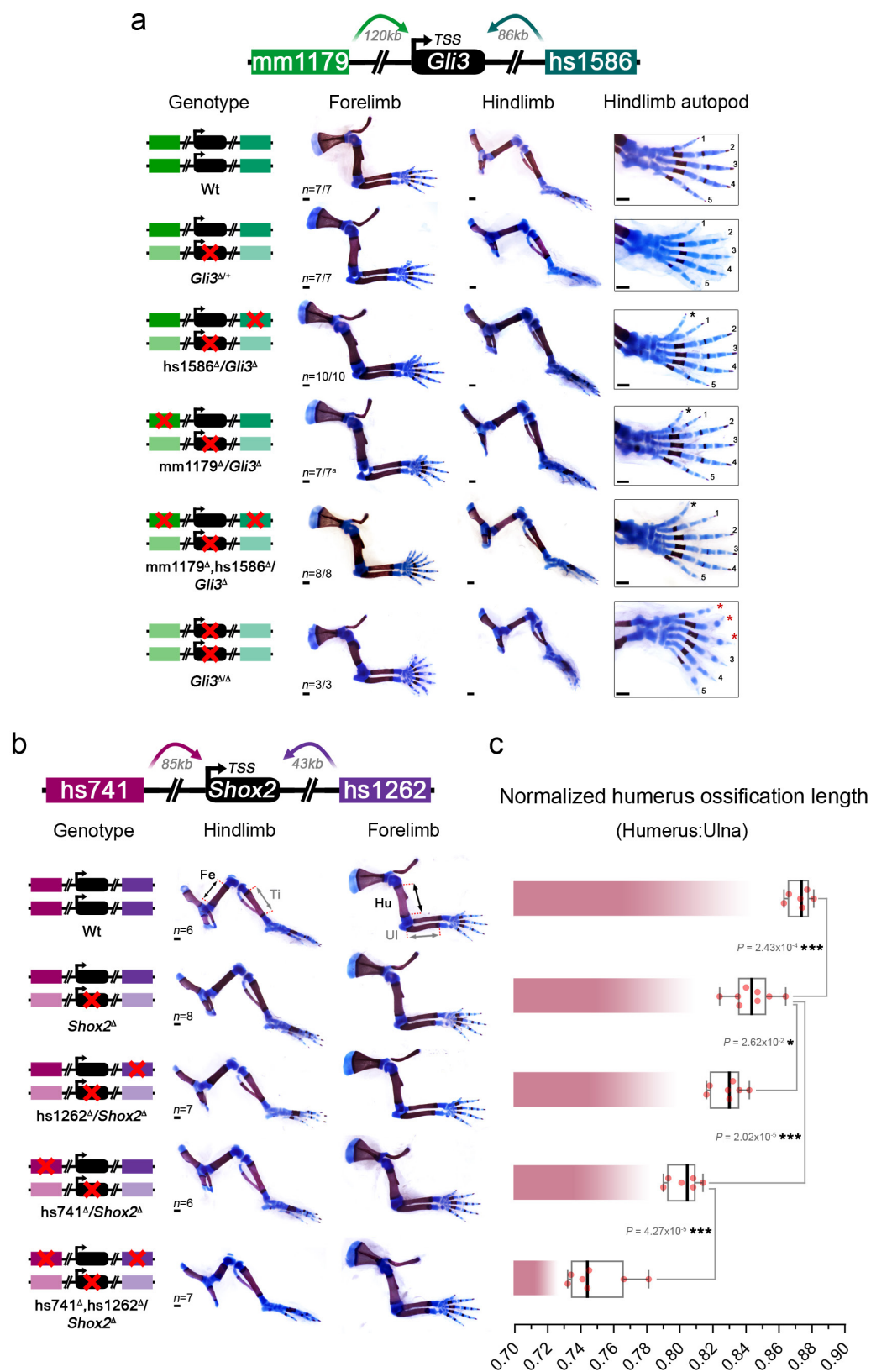
plates of double enhancer transgenic embryos. Close-ups (right) show that the anterior mesenchyme (Fig. 2c) harbours many cells with dual enhancer activities (yellow). A fraction of double enhancer-positive cells carries the signature of Sox9 digit progenitors (white, bottom).  $n = 3$  independent embryos per genotype were analysed, with similar results. Nuclei, detected via Hoechst staining, are blue. Scale bars, 100  $\mu\text{m}$  (**a, b**); 50  $\mu\text{m}$  (**c, d**).



**Extended Data Figure 7 | Generation of *Gli3* and *Shox2* knockout alleles and characterization of enhancer deletions in a sensitized background.**

**a, d**, Top, schematic showing CRISPR–Cas9-mediated deletions used to generate *Gli3* and *Shox2* loss-of-function alleles. Genotyping primers used to validate targeted deletion events are indicated. Bottom, Sanger sequencing confirmation of deletion event, with grey and red dashed lines indicating breakpoints. Right, PCR genotyping examples with the size of the product specific for the deletion allele depicted in red (primers listed in Supplementary Table 3). **b**, *In situ* hybridization showing the gradual decrease in anterior *Gli3* transcript in forelimbs of wild-type, *Gli3*<sup>Δ/+</sup> and sensitized mm1179/hs1586 DKO (DKO/*Gli3*<sup>Δ</sup>) embryos. **c**, qPCR validation of *Gli3* mRNA levels in forelimb hand plates from the genotypes

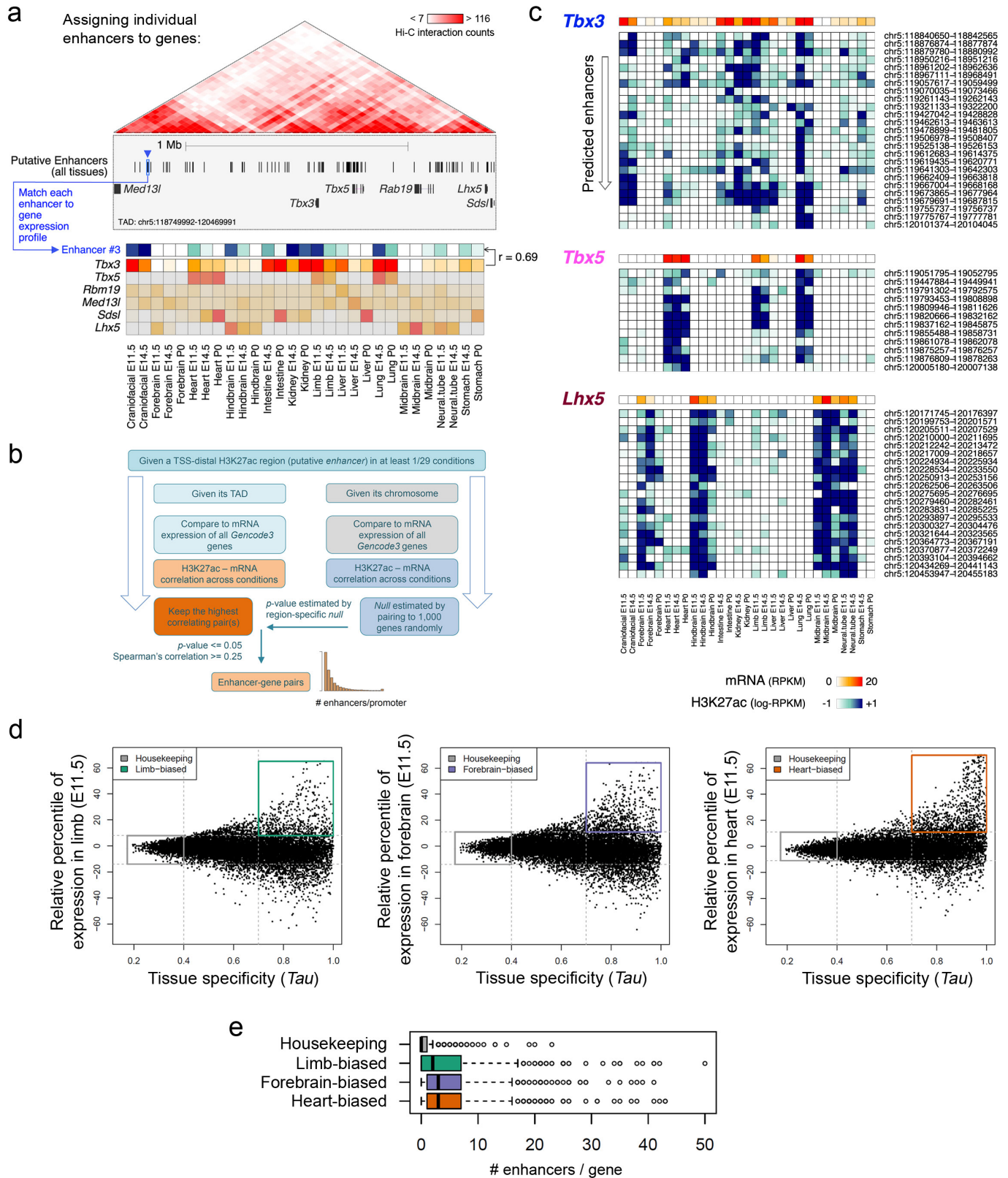
shown in **b**. **e**, *Shox2* expression (*in situ* hybridization) in forelimbs and hindlimbs of wild-type, *Shox2*<sup>Δ/+</sup> and sensitized hs741/hs1262 DKO (DKO/*Shox2*<sup>Δ</sup>) embryos. Arrowheads point to the domains where *Shox2* expression is nearly abolished in enhancer DKO/*Shox2*<sup>Δ</sup> embryos. **f**, qPCR revealing significantly downregulated *Shox2* mRNA levels in hindlimbs of DKO/*Shox2*<sup>Δ</sup> compared to *Shox2*<sup>Δ/+</sup> embryos. *n* indicates the number of independent biological replicates with similar results. Bar plots illustrate mean and s.d., with individual biological replicates shown. \*\*\**P* < 0.001; \**P* < 0.05 (two-tailed, unpaired *t*-test). n.s., not significant. For *in situ* hybridization, transcript distribution was reproduced in at least *n* = 3 independent biological replicates. Scale bars, 100 μm.



Extended Data Figure 8 | See next page for caption.

**Extended Data Figure 8 | Limb phenotypes of individual and combinatorial *Gli3* and *Shox2* enhancer knockouts in the presence of reduced target gene dosage.** **a**, Skeletal phenotypes resulting from mm1179 and hs1586 enhancer deletions in combination with reduction to one copy of the *Gli3* gene at E18.5. Genotypes are shown on the left with red crosses indicating elements deleted by CRISPR–Cas9. While forelimbs of *Gli3*<sup>Δ/+</sup> embryos displayed bifurcated digit 1 terminal phalanges<sup>65</sup>, hindlimbs showed an extra toe structure but without detectable cartilage template. Four out of seven mm1179<sup>Δ</sup>/*Gli3*<sup>Δ</sup> embryos displayed additional bifurcation of digit 2 of the right forelimb (**a**), which suggests that removal of mm1179 reduces *Gli3* levels in the anterior forelimb more than deletion of hs1586. An almost complete anterior extra toe formed in hindlimbs of embryos with single or dual enhancer deletions in the sensitized background (black asterisks). Loss of both *Gli3* copies resulted in anterior hindlimb polydactyly with altered digit identities (red asterisks)<sup>24</sup>.

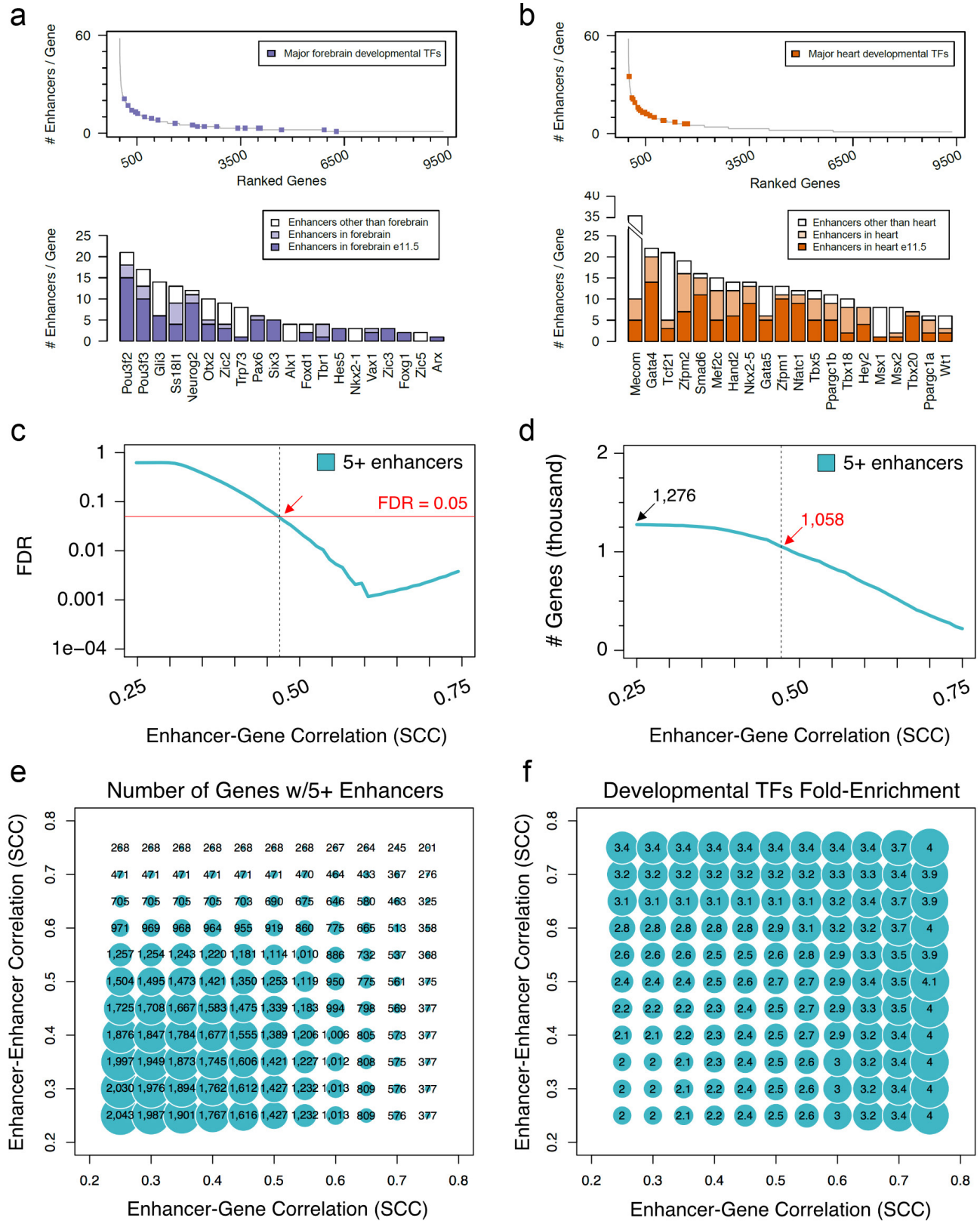
**b**, Allelic series depicting shortening of the stylopod (humerus and femur) in limb skeletons with individual or combined hs741 and hs1262 enhancer deletions in a *Shox2* sensitized condition (see also Fig. 3b). Stylopod ossification length (double arrows) appears less reduced in forelimbs (humerus, Hu) than in hindlimbs (femur, Fe) of embryos lacking the activity of both enhancers (hs741<sup>Δ</sup>, hs1262<sup>Δ</sup>/*Shox2*<sup>Δ</sup>). Tibia (Ti) and ulna (Ul) were normal in all genotypes examined. **c**, Humerus ossification length (normalized to ulna ossification length) is significantly reduced in embryos lacking either hs741 or hs1262 in the presence of only one copy of *Shox2*. In embryos lacking both enhancers in the sensitized background, significant shortening of humerus ossification is observed (compared to all other genotypes). *n* indicates the number of independent biological replicates with similar results. Box plots indicate median, interquartile values, range and individual biological replicates. \*\*\**P* < 0.001; \**P* < 0.05 (two-tailed, unpaired *t*-test). Scale bars, 500 μm.



Extended Data Figure 9 | See next page for caption.

**Extended Data Figure 9 | A correlative framework to define enhancer–promoter associations across the mouse genome.** **a**, The TAD including the transcriptional regulators *Tbx3*, *Tbx5* and *Lhx5* illustrates the statistical framework to define enhancer–promoter associations genome-wide. For each predicted enhancer, correlation between its H3K27ac signal (blue arrowhead, blue-shaded heat map) with the mRNA expression profiles of every gene in the TAD (red-shaded heat map) across all available tissues and developmental stages was assessed. The enhancer was then assigned to the most highly correlated gene, *Tbx3* in the case of enhancer 3. **b**, Schematic depicting the underlying statistical framework used to determine genome-wide enhancer–promoter interactions (see Methods). **c**, Activity pattern for the enhancers assigned to *Tbx3*, *Tbx5* and *Lhx5*. Genomic coordinates are listed on the right. For each predicted enhancer–gene pair, Spearman’s correlation coefficient (SCC,  $n = 29$ ) and the corresponding empirically estimated  $P$  value (from 1,000 random enhancer–gene pairings) are shown in Supplementary Table 11.

**d**, Identifying genes with biased expression in embryonic limb, forebrain, or heart. Expression variability across 29 RNA-seq datasets from multiple tissues and developmental time points, measures of tissue specificity (Tau ( $\tau$ ),  $x$ -axis) and specific tissue-biased expression at E11.5 ( $y$ -axis) for each protein-coding gene were calculated (see Methods). Housekeeping genes were defined as displaying  $\tau \leq 0.4$  and relative expression in the limb between the 5th and 95th percentiles. Tissue-biased genes were defined as showing  $\tau \geq 0.7$  and relative expression higher than the 95th percentile. **d**, Distribution of enhancer numbers assigned to each gene, for the different gene categories. Genes with tissue-biased expression profiles were associated with a significantly higher number of enhancers than housekeeping genes.  $P = 4 \times 10^{-121}$  ( $n = 553$ ),  $P = 7 \times 10^{-97}$  ( $n = 626$ ) and  $P = 6 \times 10^{-83}$  ( $n = 826$ ) for limb, forebrain and heart biased genes, respectively (two-sided Mann–Whitney tests).  $n = 1,287$  for housekeeping genes. Box plots indicate median, interquartile values and range. Outliers are shown as individual points.



Extended Data Figure 10 | See next page for caption.

**Extended Data Figure 10 | Enhancer redundancy as a widespread feature of developmental genes and robustness to the choice of thresholds used in the correlative approach.** **a, b**, Top, number of enhancers assigned to each gene through the correlative framework, with developmental transcription factors (TFs) showing biased expression in forebrain (**a**, blue dots) or heart (**b**, orange dots) indicated. Classification of tissue-biased developmental transcription factors is described in Methods. Genes with at least one assigned enhancer are displayed and sorted according to the number of assigned enhancers (left to right). Bottom, bar plot showing the total number of enhancers assigned to each of the transcription factors highlighted in the top panels. For each gene, a colour code shows the number of predicted enhancers assigned to that gene in the relevant tissue (**a**, heart; **b**, forebrain) at E11.5 (dark colour), in the relevant tissue at any other developmental stage included in the analysis (light colour), or in any other tissue (white). **c**, Estimated

FDR (based on genome-wide permutations, see Methods) of observing a gene with five or more enhancers assigned to it, for increasingly larger correlation coefficients (0.25 to 0.75). The red solid line indicates an FDR of 0.05. The red arrow and the black dashed line highlight the lowest correlation coefficient (0.47, considering a step of 0.01) with an  $FDR \leq 0.05$  ( $FDR = 0.0495$ ). **d**, Number of genes showing five or more enhancers assigned to them, for increasingly larger correlation coefficients (0.25 to 0.75). The total number of genes ( $SCC \geq 0.25$ ) along with the number of genes identified using the threshold set in **c** ( $SCC \geq 0.47$ ) is indicated (1,276 and 1,058, respectively; see Supplementary Tables 11, 12). **e**, Bubble plot showing the number of genes with five or more enhancers assigned to them, at increasingly higher correlation between enhancer and target gene expression ( $x$ -axis) and between enhancers assigned to the same gene ( $y$ -axis). **f**, Bubble plot displaying the fold-enrichment (linear) for developmental transcription factor genes among each set in **c**.

# Dopamine neuron activity before action initiation gates and invigorates future movements

Joaquim Alves da Silva<sup>1,2</sup>, Fatuel Tecuapetla<sup>1,3</sup>, Vitor Paixão<sup>1</sup> & Rui M. Costa<sup>1,2,4</sup>

**Deciding when and whether to move is critical for survival. Loss of dopamine neurons (DANs) of the substantia nigra pars compacta (SNc) in patients with Parkinson's disease causes deficits in movement initiation and slowness of movement<sup>1</sup>. The role of DANs in self-paced movement has mostly been attributed to their tonic activity, whereas phasic changes in DAN activity have been linked to reward prediction<sup>2,3</sup>. This model has recently been challenged by studies showing transient changes in DAN activity before or during self-paced movement initiation<sup>4–7</sup>. Nevertheless, the necessity of this activity for spontaneous movement initiation has not been demonstrated, nor has its relation to initiation versus ongoing movement been described. Here we show that a large proportion of SNc DANs, which did not overlap with reward-responsive DANs, transiently increased their activity before self-paced movement initiation in mice. This activity was not action-specific, and was related to the vigour of future movements. Inhibition of DANs when mice were immobile reduced the probability and vigour of future movements. Conversely, brief activation of DANs when mice were immobile increased the probability and vigour of future movements. Manipulations of dopamine activity after movement initiation did not affect ongoing movements. Similar findings were observed for the initiation and execution of learned action sequences. These findings causally implicate DAN activity before movement initiation in the probability and vigour of future movements.**

We quantified spontaneous motion of mice placed in an open field (with no external cues, food deprivation or reward) by using inertial sensors that measure high-resolution acceleration and angular velocity (Fig. 1a). We found that the distribution of movements in an open field was not a continuum between arrest and motion, but rather a bimodal distribution (Fig. 1b and Extended Data Fig. 1). Using the minimum acceleration between the two peaks of the bimodal distribution (Fig. 1b), we could separate periods of immobility from periods of overt mobility (Fig. 1b and Extended Data Fig. 1), and precisely identify moments of spontaneous movement initiation.

We investigated the activity of photoidentified SNc DANs in relation to movement initiation by implanting 16-channel movable electrode bundles coupled to a fibre-optic cannula placed just above the SNc (Fig. 1a). We implanted these bundles in six TH-Cre mice<sup>8</sup> (Extended Data Fig. 2a–c) crossed with Ai32 mice<sup>9</sup> to obtain expression of channelrhodopsin-2 (ChR2) in TH<sup>+</sup> neurons, and used photoidentification<sup>10</sup> to identify putative DANs (Extended Data Fig. 3, see Methods for details). Next, we built peri-event time histograms (PETH) of their activity aligned to spontaneous movement initiations. We found that the average activity of all recorded DANs increased transiently before movement initiation (Fig. 1d). Consistently, we found that many photoidentified DANs were significantly modulated by movement initiation and that the majority of these were positively modulated (Fig. 1d). The latency for modulation preceded the initiation of movement for positively modulated neurons, but not for negatively

modulated neurons (Fig. 1e). These findings were corroborated using microendoscopic calcium imaging of SNc DANs in freely moving mice (Extended Data Fig. 4; 22 neurons, see Methods and below for details).

We next examined whether the transient activity of individual DANs before action initiation was tuned to the initiation of specific actions, or represented a more general signal before action initiation. To characterize each spontaneous movement initiation, we built trajectories in the motion-sensor space using a combination of total body acceleration (correlated with displacement, see Extended Data Fig. 1), angular velocity (direction of movement) and gravitational acceleration (postural changes)<sup>11</sup> (Fig. 1f). Next, we used affinity propagation<sup>12</sup> to cluster similar movement initiations<sup>13</sup> (Fig. 1g). We found that most DANs were broadly tuned, and were transiently active before rather than different initiations (Fig. 1h). Furthermore, we found that initiation trajectories preceded by increased activity of each DAN were as variable as all other initiation trajectories (Fig. 1i). These data strongly indicate that the transient activity of individual DANs before initiation is not very action-specific.

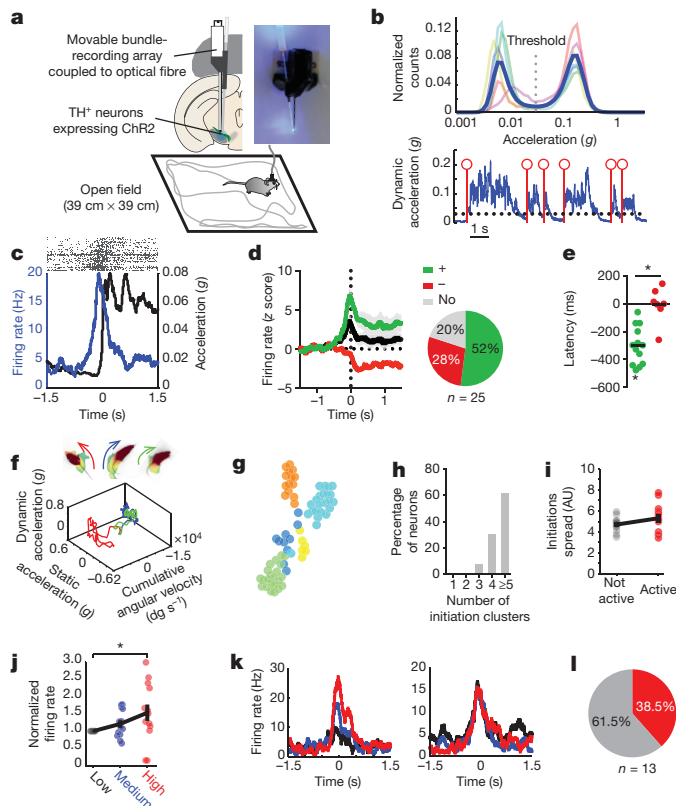
Previous studies in patients with Parkinson's disease<sup>14</sup> and animal models<sup>15</sup> have shown that dopamine depletion leads to less vigorous movements. We therefore investigated whether the activity of DANs before movement initiation encoded information about the vigour of the movement that was about to be initiated. We verified that overall, the activity of DANs 300 ms before movement initiation was significantly related to the vigour of future movements (measured by body acceleration; Fig. 1j, k). When doing per-trial analyses comparing all the lower vigour with the higher vigour initiations, we found that 38.5% of the neurons had significantly higher activity before higher vigour movements (Fig. 1l).

In order to test whether inhibiting DAN activity affects movement initiation, we expressed archaerhodopsin (ArchT)<sup>16</sup> specifically in SNc neurons (AAV2/1.CAG.Flex.ArchT-GFP injected into TH-Cre mice, Extended Data Fig. 5). We expressed ArchT in the SNc of 11 TH-Cre mice and YFP in the SNc of 9 TH-Cre mice (control group), and delivered light unpredictably for periods of 15 s (Fig. 2a). Inhibition of SNc DANs increased the probability of mice being immobile (Fig. 2b, c). This was not observed in YFP controls (Fig. 2b, c). To investigate whether DAN inhibition mainly affected movement initiation or reduced ongoing movement, we investigated the effects of DAN inhibition in trials in which the mouse was immobile (for at least 300 ms), or mobile (for at least 300 ms), when the inhibition started (Fig. 2d–h).

We found a significant impairment in movement initiation when SNc DANs were inhibited during immobility (Fig. 2g). The effect of inhibition was relatively rapid with a significant difference between light and no light after 2.4 s (Fig. 2e). Consistently, we also found that 5 s of inhibition was sufficient to impair movement initiation (Extended Data Fig. 6).

By contrast, there was no significant change in mean acceleration when inhibition happened after movement onset (Fig. 2g). There was

<sup>1</sup>Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon, Portugal. <sup>2</sup>Nova Medical School, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisbon, Portugal. <sup>3</sup>Neuropatología molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico. <sup>4</sup>Departments of Neuroscience and Neurology, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York, USA.



**Figure 1 | Movement initiation is preceded by increased activity of dopamine neurons.** **a**, Top left, photostimulation schematics. Top right, movable bundle electrode array with fibre-optic cannula. Bottom, open field set-up. The mouse brain schematic in this panel has been redrawn with permission from ref. 31. **b**, Top, distribution of total dynamic acceleration in the open field (mean in blue;  $3 \pm 2.2$  one-hour open-field sessions per mouse,  $n = 6$  mice). Bottom, examples of movement initiation events (red). Dotted line represents acceleration threshold. **c**, PETH of a positively modulated DAN aligned to movement initiation. **d**, Left, mean trace for all neurons ( $n = 25$ , black), positively modulated neurons ( $n = 13$ , green) and negatively modulated neurons ( $n = 7$ , red), aligned to movement initiation ( $93.45 \pm 35.99$  (mean  $\pm$  s.e.m.) spontaneous initiations per neuron). Grey shadows denote s.e.m. Right, proportion of positively (+) and negatively (−) modulated and unmodulated (No) neurons. **e**, Latency of each neuron to be significantly modulated in relation to movement initiation (red, negatively modulated; green, positively modulated). **f**, Time series of video frames for three movement initiations (1.5 s duration) and their representation in the motion sensor space. Cumulative angular velocity is shown in degrees per second ( $\text{deg s}^{-1}$ ). **g**, Representation of the initiations of one session using  $t$ -distributed stochastic neighbour embedding ( $t$ -SNE) dimensionality reduction. Colours represent different clusters determined using affinity propagation clustering. **h**, Number of initiation clusters in which each positively modulated neuron was significantly activated ( $n = 13$ ,  $30.4 \pm 0.19\%$  of initiations were preceded by a significant increase in neuron activity). **i**, Spread of initiations (mean distance to every other initiation) according to whether positively modulated neurons were active or not (not active,  $n = 8$ ; active,  $n = 5$ ). **j**, Normalized firing rate (firing rate / firing rate of low acceleration trials) for positively modulated neurons of low, medium and high acceleration trials ( $n = 13$ ). **k**, Examples of a vigour-related neuron (left) and a neuron unrelated to vigour (right). Red, blue and black represent firing rates during high, medium and low acceleration initiations, respectively. **l**, Proportion of vigour-related (red) and non-related neurons (grey).  $P < 0.05$ , paired  $t$ -test comparing firing rate of low and high acceleration trials. Data are mean  $\pm$  s.e.m. (**i**, **j**). \* $P < 0.05$ .

no change in the vigour of movement when SNc DANs were inhibited after movement initiation and mice did not stop during the inhibition (Fig. 2h). Furthermore, in mobile trials in which animals stopped during the 15 s ( $\geq 300$  ms immobile), there was no difference in acceleration

between laser-on and laser-off trials before the first stop. However, there was clear impairment in movement initiation after the first stop (Fig. 2i). This was not observed in YFP controls (Fig. 2j).

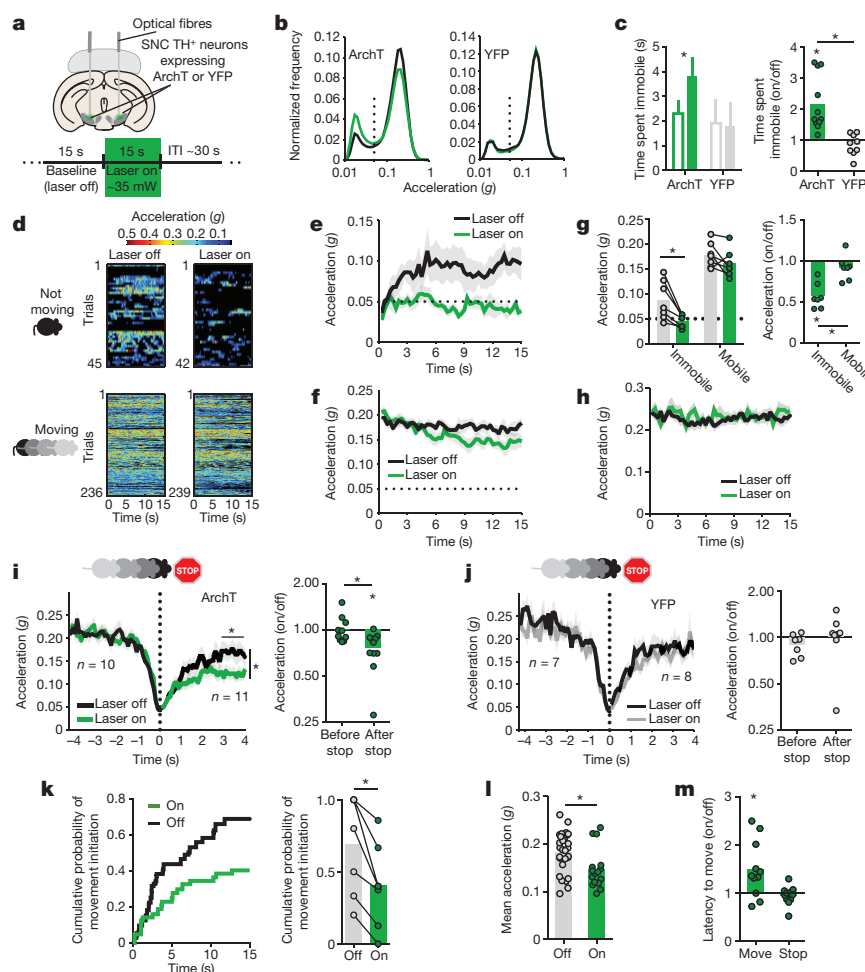
A more detailed analysis of the effects of inhibition when mice were immobile revealed that there was a significant decrease in the probability of initiating movement during the 15 s of inhibition (Fig. 2k). Furthermore, even when mice were able to initiate movements, the latency to initiate was significantly higher than in laser-off trials, and the initiated movements were less vigorous (Fig. 2l, m). Taken together, these data indicate that DAN activity before movement initiation modulates the probability and vigour of future movements, but activity of these neurons is less critical for the maintenance and vigour of ongoing movements.

Next, we investigated whether brief activation of DANs, when animals were immobile, would be sufficient to promote movement initiation. We expressed ChR2 in SNc DANs using a similar Cre-dependent strategy (DIO-ChR2-YFP in seven mice, and control DIO-eYFP in five mice). Stimulation at 20 Hz for 500 ms (Fig. 3a) delivered when mice were immobile was sufficient to produce overt movement that lasted several seconds (Fig. 3b), in accordance with previous findings<sup>7,17,18</sup>. The same activation when mice were overtly moving did not significantly affect ongoing acceleration (Fig. 3b, c).

To further corroborate this finding, we performed an online closed-loop experiment in which mice received stimulation if they were immobile for at least 900 ms (in 50% of the trials, Fig. 3d). Trials in which light was not delivered (50%) were used as within-animal control (laser-off trials). Average acceleration during the first second after the closed-loop trigger was higher during laser-on than during laser-off trials in the ChR2 group (Fig. 3e–g). Moreover, the latency to initiate movement when SNc DANs were briefly activated was almost three times shorter than in laser-off trials (Fig. 3h). The percentage of trials in which movement was initiated during the first second was also higher in laser-on trials (Fig. 3i). We found no evidence that this closed-loop SNc DAN activation had a reinforcement effect, because immobility states did not become more frequent in ChR2 mice (interval between immobility periods: ChR2,  $254.5 \pm 116.5$  s; YFP,  $150.5 \pm 65.8$  s;  $t = 1.74$ ,  $P = 0.12$ ). We found that movements initiated during laser-on trials were more vigorous than movements spontaneously initiated during laser-off trials (Fig. 3j, k). None of the described effects were found in YFP controls.

The results presented above highlight a specific role for the transient activity of DANs for the gating and invigoration of self-paced movement initiation, but not for the modulation of ongoing movements. On the basis of these findings, one prediction would be that if individual spontaneous movements are chunked into a sequence of movements, then the activity of DANs would become preferentially active before sequence initiation, but not during the execution of individual elements within the sequence. To test this prediction, we trained mice on a self-paced operant task in which eight lever presses led to a 20% sucrose solution reward (fixed-ratio eight task (FR8)), without any explicit stimuli signalling the availability of reward<sup>4</sup> (Fig. 4a). We implanted a gradient index lens just above the SNc of four TH-Cre mice (Extended Data Fig. 7d), and injected a virus that expressed GCaMP6f<sup>19</sup> in a Cre-dependent manner<sup>20</sup> (AAV2/5.CAG.Flex.GCaMP6f). We then used a miniaturized epifluorescence microscope<sup>21</sup> to image calcium transients in genetically identified SNc DANs while mice were performing the FR8 task (Fig. 4b, c). Similarly to previous findings<sup>4</sup>, by creating lever press time histograms using normalized fluorescence traces ( $z$  score of  $\Delta F$ ), we found that the proportion of modulated neurons was different between press events with the highest proportion of neurons being modulated by the first press (Fig. 4d–f). As predicted, this higher proportion of neurons that were related to the first press was not apparent early in training, and developed with sequence learning<sup>4,22</sup> (Extended Data Fig. 8).

Recent studies claimed that SNc neurons are more modulated by movement, whereas ventral tegmental area neurons are more modulated by reward<sup>7</sup>. However, we found that around 35% of neurons in the



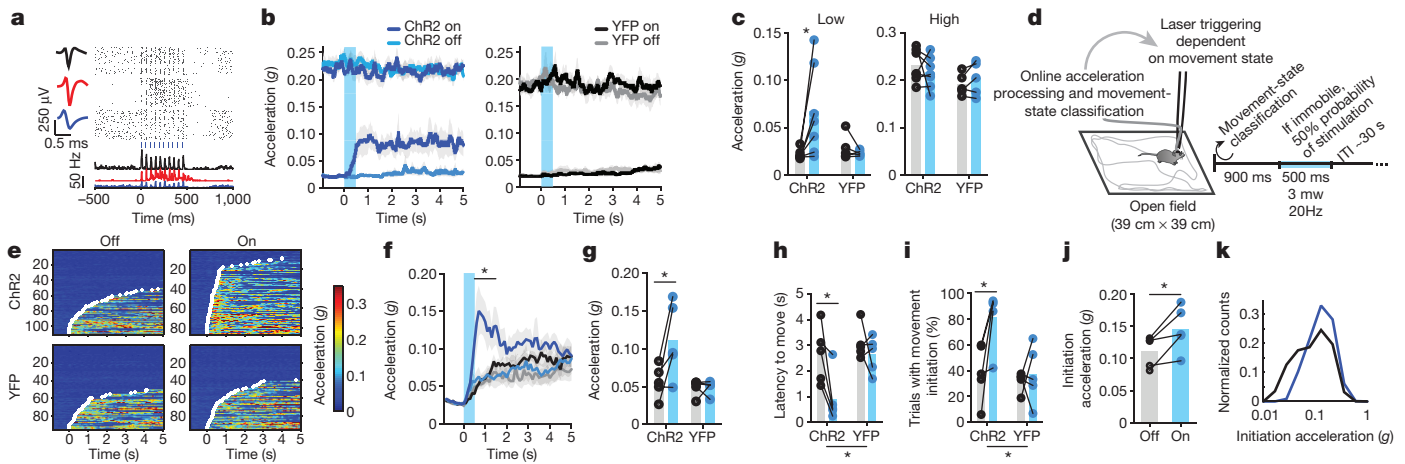
**Figure 2 | Inhibition of SNc dopamine neurons impairs movement initiation, but not ongoing movement.** **a**, Schematics showing fibre positioning and trial structure. **b**, Distribution of acceleration in the open field during laser on and laser off for ArchT (left) and YFP (right) groups. The vertical dashed line denotes the acceleration threshold. **c**, Left, time spent immobile during laser-off and laser-on periods. Clear bars indicate laser off and filled bars indicate laser on. Right, time spent immobile during laser on normalized to the baseline.  $n = 11$  ArchT mice, 9 YFP mice. Single asterisk above ArchT indicates significant difference from baseline (1). **d**, Heat maps of acceleration data of all trials where the mouse was immobile (top) or mobile (bottom) before the start of the trial, in laser-off (left) and laser-on (right) conditions.  $n = 11$  ArchT mice. **e**, **f**, Acceleration during laser-off and laser-on trials for immobile trials and mobile trials.  $n = 7$  ArchT mice. The horizontal dotted line denotes the acceleration threshold. There was an interaction between inhibition and mobility state when inhibition started (Supplementary

Table 1). **g**, Left, mean values of the data plotted in **e** and **f**. Grey, laser off; green, laser on. Right, same data normalized (laser on / laser off). Single asterisk in Immobile indicates significant difference from 1. **h**, Mean acceleration per mouse for mobile trials without stops.  $n = 10$  ArchT mice. **i**, Left, mobile trials aligned to the first stop. Right, normalized mean acceleration (laser on / laser off) before ( $-4$  to  $-3$  s) and after ( $3$ – $4$  s) the first stop for ArchT mice. Single asterisk above 'After stop' indicates significant difference from 1. **j**, Same as in **i**, but for the YFP group. **k**, Left, mean cumulative probability of movement initiation for immobile trials. Right, mean probability to initiate movement.  $n = 7$  ArchT mice. **l**, Mean acceleration for initiations that occurred during immobile trials of ArchT mice. Laser on,  $n = 16$ ; laser off,  $n = 30$  trials. **m**, Normalized mean latency to initiate movement.  $n = 9$  YFP mice, 11 ArchT mice. Single asterisk above Move indicates significant difference from 1. Data are mean  $\pm$  s.e.m. Error bars and shaded areas denote s.e.m.  $*P < 0.05$ .

SNc responded to reward (Fig. 4e). This is almost similar to the percentage of neurons modulated during sequence initiation (approximately 40% first press neurons; Fig. 4d, e). Importantly, there was little overlap between reward- and first-press-modulated neurons, and the overlap was not significantly different than what would be expected by chance (Extended Data Fig. 9). These data suggest that different populations of SNc DANs are related to movement versus reward, but it remains to be determined whether these correspond to populations that project to the dorsal versus ventral striatum, as has previously been suggested<sup>23</sup>.

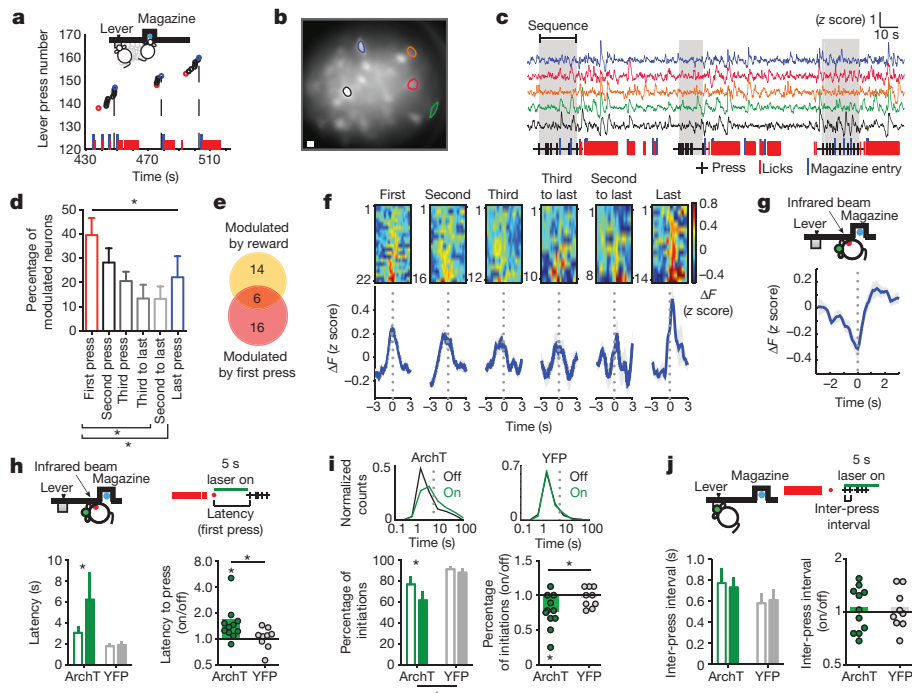
Next, we tested whether SNc DAN activity was necessary for sequence initiation. TH-Cre mice expressing ArchT ( $n = 11$ ) or YFP ( $n = 9$ ) were trained on the FR8 task for 12–14 days. Mice developed a structured behaviour with predictable sequence initiations and trajectory after reward consumption<sup>24</sup> (Fig. 4a). To inhibit SNc neurons before the first press of a sequence, the laser was triggered when

mice broke an infrared beam placed between the reward magazine and the lever (Fig. 4g, h), corresponding to the moment of minimal DAN activity (before the increase in activity of first press neurons; Fig. 4g). We compared a block of inhibition (laser-on block) with a previous block without inhibition (laser-off block) during the same session. Inhibition during 5 s before the first lever press resulted in a significant increase in the latency to initiate the action sequence when compared to laser-off trials (Fig. 4h). Moreover, the probability of initiating a sequence decreased during the 5 s of DAN inhibition (Fig. 4i). Consistent with the experiments presented above, when the inhibition happened after sequence initiation (triggered by the first press), the inter-press interval and number of presses during the 5-s inhibition were not altered (Fig. 4j). No effects were observed in YFP controls (Fig. 4h–j). These results were replicated using DAT-IRES-Cre mice using a different inhibitory opsin (Jaws<sup>25</sup>) when pseudorandomly



**Figure 3 | Transient SNc dopamine neuron activation promotes movement initiation.** **a**, Example of three SNc DANs (single units) expressing ChR2, following stimulation at 20 Hz. **b**, Mean acceleration depending on movement state before the trial.  $n = 7$  ChR2 mice;  $n = 5$  YFP mice. **c**, Mean acceleration from 0 to 1 s depending on movement state before the trial. Left, immobile; right, mobile.  $n = 7$  ChR2 mice;  $n = 5$  YFP mice. Blue, laser on; grey, laser off. **d**, Closed loop set-up. **e**, Heat maps of acceleration data for all trials. Laser-trigger criteria were reached at time 0. White crosses indicate onset of movement. **f**, Mean acceleration.  $n = 5$  mice per group. Dark blue, ChR2 on; light blue, ChR2 off; black, YFP on; grey, YFP off. **g**, Mean acceleration from 0 to 1 s.  $n = 5$  mice per group. **h**, Latency to initiate movement.  $n = 5$  mice per group. **i**, Percentage of trials with movement initiation between 0 and 1 s.  $n = 5$  mice per group. In **h**, **i**, there is a significant effect for group between ChR2 and YFP. **j**, Mean distribution of acceleration during the first second after movement initiation for laser on (blue) and laser off (black).  $n = 5$  ChR2 mice. **k**, Mean acceleration during the first second of movement initiation during laser-off and laser-on trials.  $n = 5$  ChR2 mice. Grey bars indicate laser off; blue bars indicate laser on. Data are mean  $\pm$  s.e.m. Error bars and grey-shaded areas represent s.e.m.  $*P < 0.05$ .

YFP off. **g**, Mean acceleration from 0 to 1 s.  $n = 5$  mice per group. **h**, Latency to initiate movement.  $n = 5$  mice per group. **i**, Percentage of trials with movement initiation between 0 and 1 s.  $n = 5$  mice per group. In **h**, **i**, there is a significant effect for group between ChR2 and YFP. **j**, Mean distribution of acceleration during the first second after movement initiation for laser on (blue) and laser off (black).  $n = 5$  ChR2 mice. **k**, Mean acceleration during the first second of movement initiation during laser-off and laser-on trials.  $n = 5$  ChR2 mice. Grey bars indicate laser off; blue bars indicate laser on. Data are mean  $\pm$  s.e.m. Error bars and grey-shaded areas represent s.e.m.  $*P < 0.05$ .



**Figure 4 | SNc dopamine neurons are transiently active at sequence initiation, and when inhibited, impair sequence initiation, but not sequence performance.** **a**, Example of the behaviour microstructure during the FR8 task late in training. Red, black and blue circles indicate first, middle and last press, respectively. Red and blue bars indicate licks and head entries, respectively. Dashed line denotes reward delivery. **b**, Field of view (projection of pixel standard deviation) of a TH-Cre mouse expressing GCaMP6f in the SNc. Regions of interest correspond to traces in **c**. Scale bar, 20  $\mu$ m. **c**, Example traces obtained using the CNMF-E algorithm during FR8 task. **d**, Percentage of neurons modulated by press events.  $n = 4$  mice. **e**, Venn diagram representing reward- and first press-related neurons. **f**, PETH of positively modulated neurons for each press event (bottom) and the corresponding heat maps (top). Grey shadow denotes s.e.m. **g**, Activity of first lever press-responsive neurons aligned to the cross from the magazine to the lever, before the first lever press.  $n = 22$ . **h**, Left, latency to initiate lever press sequence. Right, normalized latency to sequence initiation. ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice. Single asterisk above ArchT indicates significant difference from 1. **i**, Top, distribution of latencies to sequence initiation. Bottom left, percentage of early initiations (latency  $< 5$  s). ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice. There is a significant effect for group between ArchT and YFP. **j**, Left, press rate. Light was delivered after the first press. Right, mean press rate normalized. ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice.  $*P < 0.05$ . Clear bars indicate laser-off and filled indicate laser-on blocks in **h**–**j**. Data are mean  $\pm$  s.e.m.

denotes s.e.m. **g**, Activity of first lever press-responsive neurons aligned to the cross from the magazine to the lever, before the first lever press.  $n = 22$ . **h**, Left, latency to initiate lever press sequence. Right, normalized latency to sequence initiation. ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice. Single asterisk above ArchT indicates significant difference from 1. **i**, Top, distribution of latencies to sequence initiation. Bottom left, percentage of early initiations (latency  $< 5$  s). ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice. There is a significant effect for group between ArchT and YFP. **j**, Left, press rate. Light was delivered after the first press. Right, mean press rate normalized. ArchT,  $n = 11$  mice; YFP,  $n = 9$  mice.  $*P < 0.05$ . Clear bars indicate laser-off and filled indicate laser-on blocks in **h**–**j**. Data are mean  $\pm$  s.e.m.

inhibiting 30% of the trials (Extended Data Fig. 10). Taken together, these results indicate that SNc dopamine activity before the initiation of the action sequence modulates the probability and latency of sequence initiation, but is not critical for the execution of ongoing sequences.

Here we show that SNc DAN activity modulates self-paced movement initiation. Importantly, precisely timed and state-dependent optogenetic manipulations did not change ongoing movements, indicating a specific role for SNc DAN activity for initiation. These results were corroborated using more complex movement sequences.

It has been proposed that dopamine release in the dorsal striatum is important for the regulation of movement vigour<sup>2,14,15</sup>, but it was thought that this effect was mostly due to the ongoing tonic levels of dopamine release. Our results indicate that the activity of DANs before movement onset modulates future movement vigour. This could explain why patients with Parkinson's disease select less vigorous movements to initiate<sup>14</sup>. It is also in accordance with recent studies that have shown that activity of DAN terminals in the dorsal striatum preceded spontaneous movement initiation but did not precede and even followed acceleration bursts during ongoing movement<sup>7</sup>. Our results suggest that transient changes in dopamine can function as a fast system that acts on top of tonic release to increase the probability (and vigour) of initiating movements, presumably by modulating the excitability of striatal projection neurons<sup>24,26</sup>, which receive information about the movements that are 'planned' at that exact time via glutamatergic inputs from cortex and/or thalamus. This suggests a role for dopamine in gating and invigorating movements that were planned elsewhere<sup>27,28</sup>, and is consistent with the observation that DAN activity is not very action-specific. More sustained changes in DAN activity could represent states in which the 'gate' is more permissive, increasing the probability of action initiation during longer periods of time, therefore promoting movement<sup>29</sup>. This would translate into more movement variability with exploration of the action space, which could be important in situations of uncertainty or learning.

These results highlight that approaches aimed at providing transient modulations of basal ganglia circuitry tied to movement initiation, for example, via closed-loop deep-brain stimulation<sup>30</sup> triggered by activity in cortical areas related to motor planning, could be beneficial to patients with Parkinson's disease.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 February 2016; accepted 11 December 2017.**

**Published online 31 January 2018.**

- Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* **79**, 368–376 (2008).
- Niv, Y., Daw, N. D. & Dayan, P. How fast to work: response vigor, motivation and tonic dopamine. *Adv. Neural Inf. Process. Syst.* **18**, 1019–1026 (2005).
- Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30**, 259–288 (2007).
- Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457–462 (2010).
- Syed, E. C. J. et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat. Neurosci.* **19**, 34–36 (2016).
- Dodson, P. D. et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in Parkinsonism. *Proc. Natl Acad. Sci. USA* **113**, E2180–E2188 (2016).
- Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).
- Gong, S. et al. Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
- Madisen, L. et al. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.* **15**, 793–802 (2012).

- Lima, S. Q., Hromádka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS ONE* **4**, e6099 (2009).
- Klaus, A. et al. The spatiotemporal organization of the striatum encodes action space. *Neuron* **95**, 1171–1180 (2017).
- Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
- Van Der Maaten, L. & Hinton, G. H. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Mazzoni, P., Hristova, A. & Krakauer, J. W. Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J. Neurosci.* **27**, 7105–7116 (2007).
- Panigrahi, B. et al. Dopamine is required for the neural representation and control of movement vigor. *Cell* **162**, 1418–1430 (2015).
- Han, X. et al. A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front. Syst. Neurosci.* **5**, 18 (2011).
- Barter, J. W. et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front. Integr. Neurosci.* **9**, 39 (2015).
- Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
- Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
- Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
- Ghosh, K. K. et al. Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
- Wassum, K. M., Ostlund, S. B. & Maidment, N. T. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biol. Psychiatry* **71**, 846–854 (2012).
- Parker, N. F. et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016).
- Tecuapetla, F., Jin, X., Lima, S. Q. & Costa, R. M. Complementary contributions of striatal projection pathways to action initiation and execution. *Cell* **166**, 703–715 (2016).
- Chuong, A. S. et al. Noninvasive optical inhibition with a red-shifted microbial rhodopsin. *Nat. Neurosci.* **17**, 1123–1129 (2014).
- Kravitz, A. V. et al. Regulation of Parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).
- Wong, A. L., Lindquist, M. A., Haith, A. M. & Krakauer, J. W. Explicit knowledge enhances motor vigor and performance: motivation versus practice in sequence tasks. *J. Neurophysiol.* **114**, 219–232 (2015).
- Thura, D. & Cisek, P. The basal ganglia do not select reach targets but control the urgency of commitment. *Neuron* **95**, 1160–1170 (2017).
- Spielewoy, C. et al. Behavioural disturbances associated with hyperdopaminergia in dopamine-transporter knockout mice. *Behav. Pharmacol.* **11**, 279–290 (2000).
- Rosin, B. et al. Closed-loop deep brain stimulation is superior in ameliorating Parkinsonism. *Neuron* **72**, 370–384 (2011).
- Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. Vaz for mouse colony management, I. Vaz for the help during photoidentification experiments, L. Perry for help with stereological cell counts, A. Klaus, P. Zhou, L. Paninski for help with the application of the CNMF-E analysis, and the Champalimaud Hardware Platform (F. Carvalho, A. Silva, D. Bento) for help with the development of the motion sensors. This work was supported by fellowships from Gulbenkian Foundation to J.A.d.S. and Grants from Fundação para a Ciência e Tecnologia, Frontiers de la Ciencia-CONACYT-2022 and the IN226517 DGAPA-PAPIIT-UNAM to F.T. and from ERA-NET, European Research Council (COG 617142), and HHMI (IEC 55007415) to R.M.C.

**Author Contributions** J.A.d.S. and R.M.C. designed the experiments and analyses and wrote the paper. J.A.d.S. performed all experiments and analyses, F.T. helped with optogenetic and recording experiments, V.P. helped with accelerometer experiments and accelerometer data analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.M.C. ([rc3031@columbia.edu](mailto:rc3031@columbia.edu)).

**Reviewer Information** Nature thanks D. J. Surmeier and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Animals.** All experiments were approved by the Portuguese DGAV and Champalimaud Centre for the Unknown Ethical Committee and performed in accordance with European guidelines. TH-Cre male mice from the F112 mouse line<sup>8</sup> between 3 and 5 months, DAT IRES-Cre<sup>32</sup> and TH-Cre;Ai32<sup>9</sup> between 2.5 and 6 months were used.

**Sample sizes, randomization and blinding.** The number of animals in each experiment was based on previous studies using a power of 0.7 and  $\alpha = 0.05$ .  $n$  was larger than 5 in all experiments (except imaging experiments) as required for the use of parametric statistics. No formal method of randomization was used; littermates were equally divided among the groups that were compared. There was no blinding of experimental groups. Every experiment contained all experimental groups that were tested concomitantly. The timing of optogenetic manipulations was controlled automatically, not by the experimenter.

**Recombinant adeno-associated viral vectors.** The following Cre-dependent adeno-associated viral vectors were used in the experiments: AAV2/5.CAG.Flex.GCaMP6f.WPRE.SV40 (titre  $1.19 \times 10^{13}$ , University of Pennsylvania); AAV2/1.CAG.Flex.ArchT-GFP (titre  $1.4 \times 10^{12}$ , University of Pennsylvania); AAV2/1.ChR2-eYFP (titre  $1.4 \times 10^{13}$ , University of North Carolina), AVV2/1.EF1a.DIO.eYFP (titre  $1.4 \times 10^{13}$ , University of North Carolina); AAV8/hSyn.Flex.Jaws-GFP (titre  $4.2 \times 10^{12}$ , University of North Carolina); rAAV8/hSyn.DIO.eGFP, (titre  $4.9 \times 10^{12}$ , University of North Carolina).

**Virus injections, electrode, lens and fibre placement.** Surgeries were performed using a stereotaxic system (Kopf). Mice were kept in deep anaesthesia using a mixture of isoflurane and oxygen (1–3% isoflurane at  $1 \text{ l min}^{-1}$ ).

For imaging experiments, a  $1 \mu\text{l}$  of virus solution was injected in the right substantia nigra compacta at the following coordinates:  $-3.16 \text{ mm}$  anteroposterior,  $1.40 \text{ mm}$  lateral from Bregma and  $4.20 \text{ mm}$  deep from the brain surface. The injection was done through a glass pipette using a Nanojet II (Drummond Scientific) with a rate of injection of  $4.6 \text{ nl}$  every  $5 \text{ s}$ . After the injection was finished, the pipette was left in place for 10–15 min. The virus solution was kept at  $-80^\circ\text{C}$  and thawed at room temperature just before the injection. A  $500\text{-}\mu\text{m}$  diameter,  $8.2\text{-mm}$  long gradient index (GRIN) lens (GLP-0584, Inscopix) was implanted at the same coordinates as the injection. Before the lens was lowered, a blunt 28G needle was lowered to  $3 \text{ mm}$  deep from the brain surface to facilitate the lowering of the GRIN lens. The GRIN lens was then lowered ( $4.2 \text{ mm}$  deep). The lens was fixed in place using cyanoacrylate and black dental cement (Ortho-Jet). One  $1/16\text{-inch}$  stainless-steel screw (Antrin miniatures) was attached to the skull to provide a scaffold to build a dental-cement-based cap that protected and fixed the lens to the skull.

Three weeks after surgery, the mouse was anaesthetized and fixed with head bars. A baseplate (BPC-2, Inscopix) attached to a mini epifluorescence microscope (nVista HD, Inscopix) was positioned above the GRIN lens. To correctly position the baseplate, brain tissue was imaged through the lens to find the appropriate focal plane using 20% LED power, a frame rate of  $5 \text{ Hz}$  and a digital gain of 4. Once the focal plane was set, the baseplate was cemented to the rest of the cap using the same dental cement. Imaging started 2–3 days after this final step.

The same stereotaxic system and anaesthesia protocol was used for electrode and fibre placement. In the case of TH-Cre;Ai32 mice, no virus injection was used.

The same coordinates were used for optrode placement, except for a depth of  $3.8\text{--}3.9 \text{ mm}$  from the brain surface. The ground wire was attached to a  $1/16\text{-inch}$  stainless-steel screw (Antrin miniatures), touching the surface of the brain.

For optogenetic experiments (ChR2 and ArchT groups), the same procedure and coordinates were used, except that a  $1.5\text{-}\mu\text{l}$  virus solution was injected bilaterally at  $2.3 \text{ nl}$  every  $5 \text{ s}$  and optical fibres with a diameter of  $230 \mu\text{m}$  and a NA of 0.39 (Thorlabs FMT 200 EMT) were placed bilaterally at a depth of  $3.9 \text{ mm}$  from the brain surface. Optical fibres were built based on a published protocol<sup>33</sup>. Two TH-Cre mice underwent the same virus injection protocol as the ArchT group and were used to obtain the data presented in Extended Data Fig. 5.

For the Jaws experiment, the same procedure was followed except that  $1 \mu\text{l}$  of virus was used and optical fibres with a diameter of  $400 \mu\text{m}$  and a NA of 0.5 (Thorlabs FP400URT) were implanted at the same depth.

**Optogenetic set-ups.** For ChR2-expressing mice and corresponding controls, light from a free-launched  $200\text{-mW}$ ,  $473\text{-nm}$ , diode-pumped, solid-state laser (Laserglow Technologies), controlled using an AOM (AA Optoelectronic), was delivered after being captured by a collimator and split using a one-input to two-outputs rotary joint (Doric Lenses). In addition,  $200\text{-nm}$ ,  $0.22 \text{ NA}$  optical fibre patch cords were used to guide the light to the fibres implanted in the mice.

For the ArchT and the corresponding controls, the same set-up was used, but with a different light source (free-launched  $500\text{-mW}$ ,  $556\text{-nm}$ , diode-pumped, solid-state laser from CNI Lasers).

For the Jaws group and corresponding controls, we used a red LED (around  $100 \text{ mW}$  maximum output, approximately  $625 \text{ nm}$ , Prizmatix). The light was

captured by a large diameter optical fibre ( $1 \text{ mm}$ ), which connected to a one-input to one-output rotary joint. A branched  $500\text{-}\mu\text{m}$  optical fibre was then used to connect to the fibres that had been implanted in the mice.

Light intensity was measured before and during experiments using a fibre similar to the ones implanted and a power meter (PD1000-S130C, Thorlabs). The power was adjusted at the tip of the fibre to be around  $15 \text{ mW}$  for photoidentification experiments (Fig. 1 and Extended Data Fig. 3), approximately  $3 \text{ mW}$  for ChR2 (Fig. 3), around  $35 \text{ mW}$  for ArchT experiments (Figs 2, 4 and Extended Data Fig. 5), and approximately  $9 \text{ mW}$  in Jaws experiments (Extended Data Fig. 10).

**Open field.** We used a  $39 \times 39 \text{ cm}$  open field with black walls ( $17.5 \text{ cm}$  height) and white acrylic floor to assess the spontaneous movement of mice. The open field was inside a sound-attenuating chamber. Illumination was provided by white ( $2700 \text{ K}$ ) LEDs (Diode, Ikea) that were placed on the floor and symmetrically around the open field in a way that illumination of the open field was uniform and indirect ( $135 \text{ lx}$ ).

**FR8 operant task.** Behaviour training and testing took place in operant chambers as described previously<sup>4</sup>. In brief, each chamber ( $23 \text{ cm L} \times 20 \text{ cm W} \times 19.5 \text{ cm H}$ ) was housed within a sound-attenuating box (Med-Associates) and equipped with one retractable lever on the left side of the food magazine and a house light ( $3 \text{ W}$ ,  $24 \text{ V}$ ) mounted on the left lateral wall. Sucrose solution ( $10\%$ ) was delivered into a metal cup in the magazine through a syringe pump ( $20 \mu\text{l}$  per reward). Magazine entries were recorded using an infrared beam and licks using a contact lickometer. Mice were placed on food restriction throughout training, and fed daily after the training sessions with approximately  $2 \text{ g}$  of regular food to allow them to maintain a body weight of around  $85\%$  of their baseline weight.

Training started with a 30-min magazine training session, in which the reinforcer was delivered on a random time schedule, on average every  $60 \text{ s}$  (30 reinforcers). The following day lever-pressing training started with continuous reinforcement (CRF), in which animals obtained a reinforcer after each lever press. The session began with the illumination of the house light and insertion of the lever, and ended with the retraction of the lever and by turning off the house light. On the first day of CRF, the sessions lasted  $45 \text{ min}$  or until mice received five reinforcers, the second day of CRF lasted  $45 \text{ min}$  or until mice received 15 reinforcers, and the last day of CRF lasted  $45 \text{ min}$  or until mice received 30 reinforcers. This last CRF session was repeated if mice failed to obtain 30 rewards within the time limit. After CRF, animals started to be trained (day 1) on a fixed ratio schedule in which eight presses earn a reinforcer (FR8), without any stimulus signalling when eight presses were completed or when the reinforcer was delivered; this training continued for 12–14 days.

All timestamps of lever presses, magazine entries and licks for each animal were recorded with a  $10\text{-ms}$  resolution. The same training chamber was used during imaging and optogenetic experiments.

**Acceleration and video recordings.** In the experiments where photoidentified TH<sup>+</sup> neurons were recorded, acceleration was recorded using a digital 9-axis inertial sensor with a sampling rate of  $200 \text{ Hz}$  (MPU-9150, InvenSense) assembled on a custom-made PCB and connected to a computer via a custom-made USB interface PCB (Champalimaud Foundation Hardware Platform).

For the other experiments, an analogue 3-axis inertial sensor was used with a sampling rate of  $1,000 \text{ Hz}$  (LIS331AL, ST) assembled on a custom PCB (Champalimaud Foundation Hardware Platform), and the signals were fed to the analogue inputs of a Cerebus recording system (Blackrock Microsystems).

Acceleration data obtained from these sensors aggregates acceleration from two sources: acceleration generated from gravity and from the body. To separate these two components we processed data obtained from both types of inertial sensor using a custom MATLAB code. We used a standard approach<sup>34</sup> that relies on filtering the 3-axis data using a Butterworth filter. In our analysis we used a median filter (7 bins wide) to remove noise peaks. Then we used a  $1\text{-Hz}$  high-pass fifth-order Butterworth filter to separate the static (gravitational) component of the signal. Unless stated otherwise, we used the sum of the three vectors of acceleration as a global measurement of body acceleration for our analysis. The dynamic (body acceleration) component accurately tracked animal movement, and correlated well with pixel change in video measurements (Extended Data Fig. 1b–d).

Video recordings were obtained using a charge-coupled device camera (DFK 31BF03, Imaging Source) and a custom-developed software in Labview (National Instruments) at a rate of  $15 \text{ frames per second}$ . This software allowed us to introduce signals to the video frames to sync acceleration, neural recordings and light delivery periods.

**Classification of movement state.** We used the average distribution of acceleration in the open field for each experimental group to define the movement state of the mice. This was possible, because the average distribution of the logarithm of total body acceleration was clearly bimodal, with a very low acceleration distribution corresponding to immobility periods (with the possible exception of small and slow postural adjustments) and a high acceleration distribution corresponding

to periods of mobility (see Fig. 1b and Extended Data Fig. 1 for a comparison between video and motion-sensor acceleration data). The acceleration threshold was defined as the lowest acceleration value between the two distributions. Unless stated otherwise in the methods, initiation events were defined as transitions between periods of at least 300 ms below the threshold followed by at least 300 ms above the threshold.

**Extracellular recordings of photoidentified SNc DANs.** We used a single-drive movable microbundle (sixteen 23- $\mu$ m tungsten electrodes) with an optic-fibre guide cannula (Innovative Neurophysiology). An optical fibre with 230  $\mu$ m diameter and a NA of 0.39 (Thorlabs FMT 200 EMT) was inserted in the cannula just to the top of the electrode microbundle cannula (see Fig. 1a for schematics). The neural activity and the timestamps from the light stimulation were recorded using a Cerebus recording system (Blackrock Microsystems).

Experiments were started one week after electrode placement. Every day, we sorted putative units using an online sorting algorithm (Central Software, Blackrock Microsystems) while the mouse was in its home cage. If putative single units were isolated, we delivered a screening protocol consisting of a train of 100 blue light pulses with a 10-ms width delivered at 1 Hz. Using neurophysiology data analysis software (NeuroExplorer V4), we built PETH that were aligned to the train pulses. If any of the isolated units appeared to be modulated by the light train, the mouse was introduced to the open field and neurons were recorded for 1 h. The stimulation protocol was run again at the end of the open field session for confirmation. At the end of the experiment, the microbundle was advanced 50  $\mu$ m to record next day. We used six mice in these experiments. During this experiment, microbundles were moved on average  $433 \pm 93.1 \mu$ m.

Units were resorted using an offline sorting algorithm (Offline Sorter V3, Plexon Inc.) to isolate single units on the basis of waveform characteristics, interspike intervals and clustering. Single units together with the timestamps of the light stimulation provided by a pulse generator (Master 8, AMPI) were exported to MATLAB for analysis.

**Criteria used to photoidentify DANs.** Neural activity referenced to light pulse onset was averaged in 1-ms bins, and averaged across trials to construct a PETH, which was the basis for analysing amplitude and latency of light-related firing activity. Distributions of the PETH from  $-900$  to  $-10$  ms before light onset were considered baseline activity. We then determined which bins, slid in 1-ms steps during an epoch spanning from light onset to 50 ms after, met the criteria for significant firing rate increases. A significant increase in firing rate was defined as at least four consecutive bins had a firing rate larger than a threshold of five standard deviations above baseline activity. The latency in modulation of photoidentification was defined as the time between light onset and the first significant bin. On the basis of the distribution of latencies of significantly modulated neurons and in accordance with previous studies<sup>4,35</sup>, we used a very short latency ( $\leq 7$  ms) for neural response to light, combined with at least a 30% increase in firing rate during the light pulse, and a high correlation coefficient between spike waveforms during light on and off ( $>0.9$ ) as criteria for positive photoidentification (Extended Data Fig. 3).

**Criteria to identify neurons modulated by movement initiation.** We built a PETH for each photoidentified single unit spanning from 1,500 ms before and after movement-initiation events. Neural activity was averaged in 100-ms bins, shifted by 1 ms (100 bins, centred on current bin). Distributions of the PETH from  $-1,000$  to  $-500$  ms before light onset were considered baseline activity. We then determined which bins, slid in 1-ms steps during an epoch spanning from  $-500$  ms to 500 ms after movement initiation, met the criteria for significant firing rate changes. A significant change in firing rate was defined when at least 50 consecutive bins had a firing rate higher or lower than a threshold of 2.56 standard deviations above or below baseline activity (99% confidence interval). The latency to modulation was defined as the time between movement onset and the first of the 50 consecutive significant bins.

**Area under the receiver operating characteristic curve (auROC) analysis.** For this analysis, we used a method similar to the one described previously<sup>35</sup>. We convolved the spike trains with a function similar to a post-synaptic potential<sup>36</sup>. To produce the ROC curves, we compared the firing rate of each 50-ms bin to the firing rates during baseline ( $-1,500$  to  $-1,000$  ms before movement initiation) across trials. The auROC for each bin was calculated using trapezoidal numerical integration.

**Identification of neuron types using affinity propagation clustering.** We used the affinity propagation algorithm<sup>12</sup> to look for subtypes of significantly modulated neurons (Extended Data Fig. 3f). It is an efficient clustering algorithm that takes as inputs the similarities between pairs of observations in the dataset (in this case, the auROC traces for each modulated neuron), and finds exemplars and the clusters around them by exchanging real-valued messages between data points. We used the MATLAB (Mathworks) function made available by the authors of ref. 12 at

<http://www.psi.toronto.edu/index.php?q=affinity%20propagation>. We used the correlation between neuron auROC traces as the measure of similarity used by the algorithm. We also used  $\text{maxits} = 1,000$ ,  $\text{convits} = 100$ ;  $\text{lam} = 0.9$  and a preference equal to the median similarity of the dataset.

**Trial-by-trial analysis of significant increases in neuron activity before movement initiation.** We defined a baseline distribution of the number of spikes per 200-ms bin for a period of 1 s (from 2,200 ms before movement initiation to 1,200 ms before movement initiation) for all trials. On the basis of this distribution, we determined a criterion for each neuron such that 90% of the 200-ms bins in the baseline had a lower number of spikes than the criteria. Trials were considered to present a significant increase in neuron activity when the number of spikes during the 200-ms bin immediately preceding movement initiation was above this criterion.

**Definition of movement initiation clusters and spread.** We used a methodology previously developed by our group to use motion-sensor data to classify behaviour<sup>11</sup>.

We defined 1-s movement initiation trajectories specified by three motion-sensor variables: total body acceleration, the angular velocity of the axis most parallel to the dorsal-ventral axis of the mice and the gravitational acceleration of the same axis. We divided these trajectories in two 500-ms bins and for each bin we defined a vector composed by a concatenation of a normalized histogram for each motion-sensor variable. In the end, each trajectory was defined by the concatenation of two vectors, the first representing the distributions of the motion-sensor variables for the first 500 ms of initiation, and the second for the last 500 ms. Consequently, the final initiation vector was composed by six different motion-sensor variable distributions. We then determined a matrix representing the distances from each initiation vector to every other vector. To calculate the distances, the motion-sensor variable distribution of one vector was compared to the same motion-sensor variable distribution of another vector (range of 0–1; with 0 indicating that the vectors are exactly the same and 1 indicating that the vectors are maximally different). The differences between each motion-sensor variable distribution were then squared and summed together to find the distance between the two initiation vectors that were evaluated (range 0–6). The final distance matrix was built by comparing each initiation vector in this way and squaring the final result, obtaining a matrix that varied from 0 (exactly the same) to 36 (maximally different). The spread values presented in Fig. 1i were determined by averaging the distances between movement initiations.

After determining an initiation distance matrix for each session, we used affinity propagation<sup>12</sup> to find clusters of initiations (see the description above regarding this methodology). We provided the affinity clustering algorithm with the additive inverse of the distance matrix as a similarity matrix. We also used  $\text{maxits} = 1,000$ ,  $\text{convits} = 100$ ;  $\text{lam} = 0.9$ . To make sure that we were selecting a consistent structure for the data, we used a value for preference between the minimum and the maximum similarity of each similarity matrix that provided the highest but at the same time most stable number of clusters (that is, a value was chosen in the middle of an interval of consecutive values that provided the same number of clusters).

**t-Distributed stochastic neighbour embedding for visualization of initiations.** We used *t*-distributed stochastic neighbour embedding (*t*-SNE)<sup>13</sup> to visualize and assess the existence of structure within the movement initiations of each session (see the example in Fig. 1h). We implemented this using a MATLAB code provided by the authors of ref. 13 (<https://lvdmaaten.github.io/tsne/>). We used a 2D *t*-SNE using a perplexity of 15 to produce the image in Fig. 1h.

**DAN activity and movement vigour analysis.** We used the mean acceleration during the first 500 ms of each spontaneous initiation as a measurement of movement vigour, and separated trials into low acceleration trials (lower tertile), medium and high acceleration trials (upper tertile). We then calculated the activity of positively modulated neurons during 300 ms before movement initiation for each acceleration tertile. A neuron was considered to be vigour related if the activity during the lower tertile trials was significantly lower than the activity during the upper tertile trials.

**GCaMP6f imaging using a mini-epifluorescence microscope.** Mice were briefly anaesthetized using a mixture of isoflurane and oxygen (1% isoflurane at  $1 \text{ l min}^{-1}$ ) and the mini-epifluorescence microscope was attached to the baseplate. This was followed by a period of 20–30 min of recovery in the home cage before experiments started. Fluorescence images were acquired at 10 Hz and the LED power was set 10–20% (0.1–0.2 mW) with a gain of 4. Image acquisition parameters were set to the same values between sessions to be able to compare the activity recorded. Three GCaMP6f-expressing TH-Cre mice were imaged while freely exploring an open field. The same mice and one more were also imaged during the FR8 task. Data shown in Fig. 4 were obtained in two consecutive late training sessions (between days 7 and 13).

**Calcium image processing and analysis.** *GCaMP6f image processing and fluorescence trace extraction.* All fluorescence movies were initially processed using the Mosaic software (v.1.1.2, Inscopix). First, all frames were spatially binned by a factor of 4. To correct the movie for translational movements and rotations, the frames were registered to a reference image consisting of an average of the raw fluorescence movie. This was achieved by implementing the TurboReg registration engine<sup>37</sup> within the mosaic software. The movie was cropped after registration to remove the post-registration black borders.

*GCaMP6f fluorescence trace extraction.* Although calcium imaging using miniscopes enables researchers to image neurons in freely moving mice, it is a challenge to adequately extract neuronal signals without background contamination. Because of this, we implemented the 'constrained non-negative matrix factorization for endoscopic data' (CNMF-E) framework<sup>11,38</sup>. This recently described framework is an adaptation of the CNMF algorithm<sup>39</sup>. It can reliably deal with the large fluctuating background from multiple sources in the data, enabling accurate source extraction of cellular signals. It includes four steps: (1) initialize spatial and temporal components of single neurons without the direct estimation of the background; (2) estimate the background given the estimated spatiotemporal activity of the neurons; (3) update the spatial and temporal components of all neurons while fixing the estimated background fluctuations; (4) iteratively repeat step 2 and 3.

**Criteria to identify DANs modulated by movement initiation using GCaMP6f imaging.** We built a PETH for each neuronal trace spanning from 3 s before to 3 s after movement-initiation events. For this analysis, we considered movement initiations as transitions between a period of at least 500 ms below to a period of at least 500 ms above the acceleration threshold. Distributions of the PETH from  $-3$  to  $-1$  s before movement onset were considered baseline activity. We then determined which bins, during an epoch spanning from  $-0.5$  before to  $0.5$  s after movement initiation, met the criteria for significant  $\Delta F$  changes. A significant change in  $\Delta F$  was defined if at least 2 consecutive bins had  $\Delta F$  higher or lower than a threshold of 99% above or below baseline  $\Delta F$ .

**Criteria to identify lever-press-related and reward-related DANs using GCaMP6f imaging.** We constructed a PETH for each neuron trace spanning from  $-3$  to  $3$  s from lever press onset for the first, second, third, third to final, second to final and final press, and also for the first lick of reward. Distributions of the PETH from  $-5$  to  $-3$  s before first lever press were considered baseline activity for all press-related activity and distributions from  $-5$  to  $-3$  s of the first lick of reward PETH was considered as baseline for reward-related activity. We then searched each PETH during an epoch spanning from  $-0.5$  s to  $0.5$  s for bins that were significantly different from the baseline. A significant change in fluorescence was defined as at least three consecutive bins with fluorescence higher than a threshold of 99% above baseline  $\Delta F$ .

**Extracellular recordings during SNc DAN inhibition.** Optic fibres and electrodes were positioned  $300\mu\text{m}$  above the SNc. Light intensity was around  $35\text{ mW}$  at the tip of the fibre, corresponding to an estimated<sup>40</sup> irradiance of  $70\text{--}206\text{ mW mm}^{-2}$  at SNc depth ( $200$  to  $400\mu\text{m}$  from the fibre tip). Neural activity was recorded daily and the electrodes were moved  $50\mu\text{m}$  at the end of each recording session. We were thus able to record neural activity from different depths ( $-3.90\text{ mm}$  to  $-4.60\text{ mm}$  from the brain surface), with neurons being recorded above, within and below the SNc. We recorded from 140 units and observed that at depths where the SNc is located, more than 60% of recorded units were inhibited (Extended Data Fig. 5), whereas above and below the SNc, very few neurons were modulated. Furthermore, we only observed one single unit that was modulated by light at the depth closest to the fibre where light intensity is higher ( $-3.9$  to  $-3.95\text{ mm}$ , 0.7% of all units recorded, 7.7% of all units recorded at this depth), indicating that light delivery per se was not sufficient to change neural activity at this power. The same positioning of the fibres and light intensity was used during the open field and operant task inhibition experiments.

We used the same set-up and methodology described for the extracellular recordings of photoidentified SNc DANs, but instead of using blue light, we used green light (see 'Optogenetics set-ups'). The mean aROC traces presented in Extended Data Fig. 5a, b were calculated as described for the photoidentification experiments. The anatomical scheme depicted in Extended Data Fig. 5b was based on the histological determined position of the electrode cannula and the amount of electrode travel at each recording session.

**Open field.** Mice were introduced to the open field and green light was delivered continuously for periods of 15 s (mean of  $22 \pm 3$  trials with a mean inter-trial interval (ITI) of  $64 \pm 28$  s per mouse in the ArchT group and a mean of  $23 \pm 4$  trials with a mean ITI of  $65 \pm 33$  s per mouse in the YFP group). An open-field session was done on the day before to habituate mice to the open field and to light delivery (with similar trial structure except for light duration, which was 5 s). Data regarding this first session are shown in Extended Data Fig. 6.

**Operant task.** The same groups of ArchT and YFP mice used in the open-field inhibition experiment were trained to perform the FR8 task. Optogenetic

experiments were started at the end of the FR8 training. We used two different light-delivery schedules: continuous light delivery for 5 s before the first lever press in a sequence; and continuous light delivery for 5 s after the first lever press in a sequence. For the first condition, we made the triggering of the light contingent on the breaking of an infrared beam (IRB) positioned right next to the magazine, on the side of the lever. This way, mice coming from consuming the reward would break the IRB before they started the next sequence. Sessions were divided into two blocks: a first block of 15 trials without light delivery; and a second block of 10 trials with light delivery. The last 10 trials with no light delivery were used to compare with the 10 trials with light delivery. In a few trials, the mouse failed to break the IRB before starting the action sequence. These trials were discarded and not included in the analysis. Sessions with no light delivery were interspersed between sessions with light delivery and a first session with light delivery was only used to habituate the mice to the delivery of light and it was not analysed. The same methodology was used in Jaws experiments with the exception that instead of light on and light off blocks there was a 30% probability of switching on the laser for each trial and data were collected in three consecutive sessions for each experimental condition (inhibition before initiation and inhibition after first press).

**SNc DAN activation. Open field.** Mice were introduced to the open field and a train of blue laser light (10 ms pulses at 20 Hz, during 0.5 s) was delivered with a variable interval: after 90 s there was a 33% probability that the light was delivered and this was repeated every 10 s until light was delivered.

**Closed loop experiment in the open field.** For closed loop experiments, the same light pulse train was used, but it was delivered depending on the acceleration state of the mouse in the following way: acceleration of mice was monitored online by feeding the analogue accelerometer data through a Cerebus recording system (Blackrock Microsystems) into MATLAB (Mathworks) using Blackrock's MATLAB interface (CBMEX). Using a custom MATLAB code, we processed accelerometer data as described above. We monitored acceleration of mice using bins of 300 ms and when mice reached 900 ms below the threshold used to identify immobility, light was delivered with a 50% probability. There was a minimum of 30 s between trials. In this experiment, we considered the maximum acceleration during the first second after each movement initiation as a measurement of initiation vigour.

**Anatomical verification.** Animals were euthanized after completion of the behavioural tests. First, animals were anaesthetized with isoflurane, followed by intraperitoneal injection of ketamine–xylazine (around  $5\text{ mg kg}^{-1}$  xylazine;  $100\text{ mg kg}^{-1}$  ketamine). Animals were then perfused with  $1\times$  phosphate-buffered saline (PBS) and 4% paraformaldehyde, and brains were extracted for histological processing. Brains were kept in 4% paraformaldehyde overnight and then transferred to  $1\times$  PBS solution. Brains were sectioned coronally in  $50\text{-}\mu\text{m}$  slices (using a Leica vibratome (VT1000S) and kept in PBS solution before mounting or immunostaining experiments). Images were taken using a wide-field fluorescence microscope (Zeiss AxioImager) and the tip of the longest track found was used to determine the anatomical location of the implants (lenses, fibres and electrodes), which was represented in the corresponding Allen Brain Atlas<sup>41</sup> slice (Extended Data Fig. 7).

To estimate the specificity of the TH-Cre line, we crossed TH-Cre mice with ROSA26-eGFP mice. TH-Cre;ROSA26-eGFP mice express GFP in neurons expressing Cre recombinase. We used slices from the brain of one TH-Cre;ROSA26-eGFP mouse and used a tyrosine hydroxylase antibody (ImmunoStar) to label the ventral tegmental area (VTA) and SNc DANs (Extended Data Fig. 2). We imaged the VTA and SNc in three different slices (approximately  $-2.9\text{ mm}$ ,  $-3.3\text{ mm}$ ,  $-3.8\text{ mm}$  from Bregma) using a confocal microscope equipped with a Diode 405-nm, Argon multi-line 458–488–514-nm and DPSS 561-nm lasers (LSM710, Zeiss). We acquired  $z$  stacks ( $354\mu\text{m} \times 354\mu\text{m} \times 5\mu\text{m}$ ;  $5\text{-}\mu\text{m}$  interslice interval) in a tile that covered the VTA, SNc and areas  $200\mu\text{m}$  above and below the SNc. We imported these images into the Stereo Investigator software (MBF Bioscience) and used a stereological approach to count labelled cells ( $\text{TH}^+$  and  $\text{Cre}^+$ ) and evaluate co-localization, using a  $100\mu\text{m} \times 100\mu\text{m}$  counting frame (Extended Data Fig. 2a–c).

We also used a stereological approach to estimate the rate and specificity of infection of the AAV2/1.CAG.Flex.ArchT-GFP and the density of  $\text{TH}^+$  neurons in the SNc of ArchT and YFP TH-Cre mice (ArchT,  $n = 2$ ; YFP,  $n = 2$ ). We imaged every six slices in which the VTA and/or SNc were found, using the confocal microscope described above (three slices per mouse). Using a  $40\times$  magnification, we acquired  $z$  stacks ( $354\mu\text{m} \times 354\mu\text{m} \times 5\mu\text{m}$ ;  $5\text{-}\mu\text{m}$  interslice interval) in a tile that covered the whole SNc. These tile  $z$  stacks were imported into the stereo investigator software (MBF Bioscience) and quantification of the  $\text{TH}^+$ ,  $\text{eYFP}^+$  or  $\text{GFP}^+$  cells was performed using a  $100\mu\text{m} \times 100\mu\text{m}$  counting frame (Extended Data Fig. 2d–h).

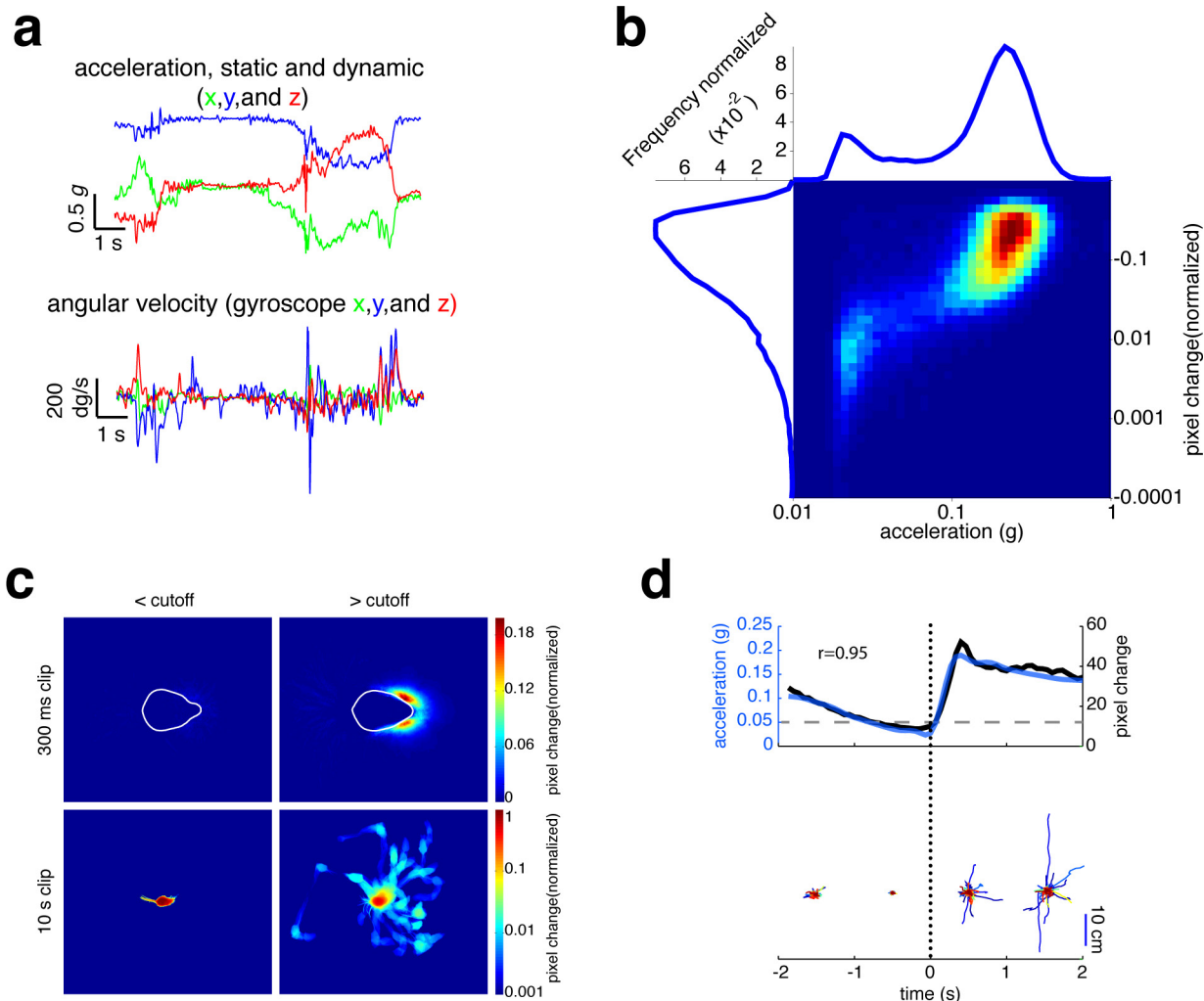
**Statistics.** Statistical hypothesis testing was done at a 0.05 significance level (except for classification of neurons, which was done at a 0.01 significance level as explained above). Parametric testing was used whenever possible to test differences

between two or more means. Normality was tested using the Shapiro–Wilk test, whereas *F*-tests (for unpaired *t*-tests) and Levene's tests (for ANOVA) were used to assess equality of variance. If data were not normally distributed, we first tried to transform the data using the natural logarithm. If the distribution of the transformed data was still not normally distributed or there was a significant difference in variance, an alternative non-parametric test was used. ANOVA and linear mixed models were used to check for main effects and interactions in experiments with repeated measures and more than one factor. The assumptions for linear mixed models were checked by careful inspection of the model residuals to check for normality and equality of variances. When main effects or interactions were significant, we did planned comparisons according to experimental design (for example, comparing laser on and off). Fisher's least significant difference tests were used for comparisons after ANOVA tests and least square means tests were used for comparison when linear mixed models were significant. Details on the statistical analysis used for hypothesis testing in the main figures can be found in the Supplementary Table 1. Statistical tests were done using Prism (GraphPad), MATLAB (MathWorks) statistical toolbox and R (R core team 2015, v.3.1.3, lme4<sup>42</sup>).

**Code availability.** MATLAB (MathWorks) codes used for data analysis are available from the corresponding author.

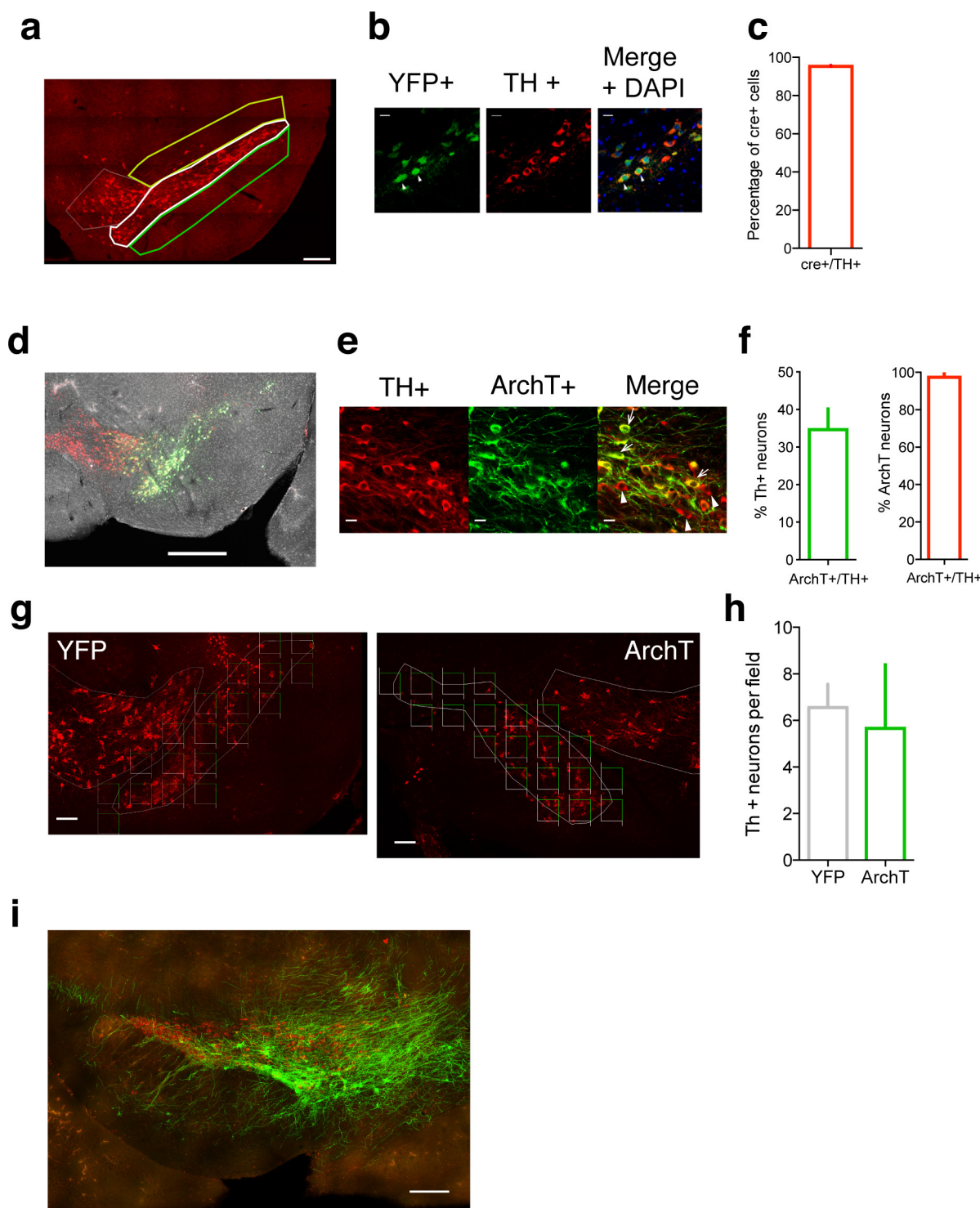
**Data availability.** Source Data for Figs 1–4 have been provided with the online version of the paper and all other data that support the findings of this study are available from the corresponding author upon reasonable request.

32. Bäckman, C. M. *et al.* Characterization of a mouse strain expressing Cre recombinase from the 3' untranslated region of the dopamine transporter locus. *Genesis* **44**, 383–390 (2006).
33. Sparta, D. R. *et al.* Construction of implantable optical fibers for long-term optogenetic manipulation of neural circuits. *Nat. Protoc.* **7**, 12–23 (2011).
34. Mathie, M. J., Lovell, N. H., Coster, A. C. F. & Celler, B. G. Determining activity using a triaxial accelerometer. In *Proc. 2nd Joint EMBS-BMES Conf.* **3**, 2481–2482 (2002).
35. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
36. Thompson, K. G., Hanes, D. P., Bichot, N. P. & Schall, J. D. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *J. Neurophysiol.* **76**, 4040–4055 (1996).
37. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
38. Zhou, P., Resendez, S. L., Stuber, G. D., Kass, R. E. & Paninski, L. Efficient and accurate extraction of *in vivo* calcium signals from microendoscopic video data. Preprint at <https://arxiv.org/abs/1605.07266> (2016).
39. Pnevmatikakis, E. A. *et al.* Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).
40. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
41. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
42. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).



**Extended Data Figure 1 | Comparison between video-motion analysis and motion-sensor data in the open field.** **a**, Example of raw acceleration (static and dynamic acceleration) and angular velocity collected using a six-axis inertial sensor. **b**, Bivariate histogram of log pixel change and log acceleration in the open field ( $n = 13$  sessions, from three mice). Notice the two clusters that emerge from this bivariate histogram (low acceleration and low pixel change cluster on the left, and high acceleration and high pixel change cluster on the right). The acceleration histogram provides a clear distinction between these two clusters. There is a high correlation between video-tracking measurements and acceleration ( $r = 0.74$ ,  $n = 62,984$  frames,  $P < 0.05$ ). This also shows that animals rely on constant acceleration from their limbs to move, and that locomotion

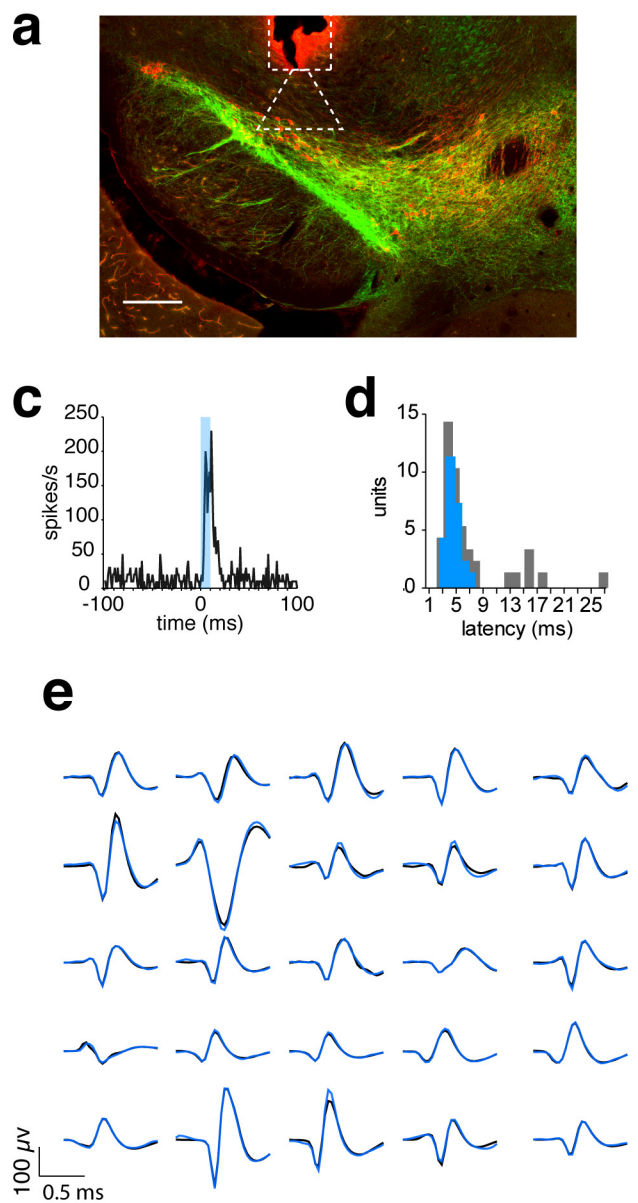
at a constant speed (acceleration 0) is highly unlikely. **c**, Heat maps of mean pixel change of video clips of 300 ms and 10 s (top and bottom, respectively), during which mice showed acceleration that was lower (left) or higher (right) than the threshold used to define immobility. **d**, Top, comparison between a video-derived motion measurement (pixel change) and total dynamic acceleration aligned to movement initiation determined using the acceleration threshold ( $n = 454$  initiations obtained from three mice during a total of 13 open-field sessions). Bottom, representation of the movement of mice (based on the centre-of-mass) during each trial within the time intervals as indicated on the  $x$  axis. The trajectories were aligned to the centre-of-mass of the last frame of each 1-s interval. Different colours denote individual trials.



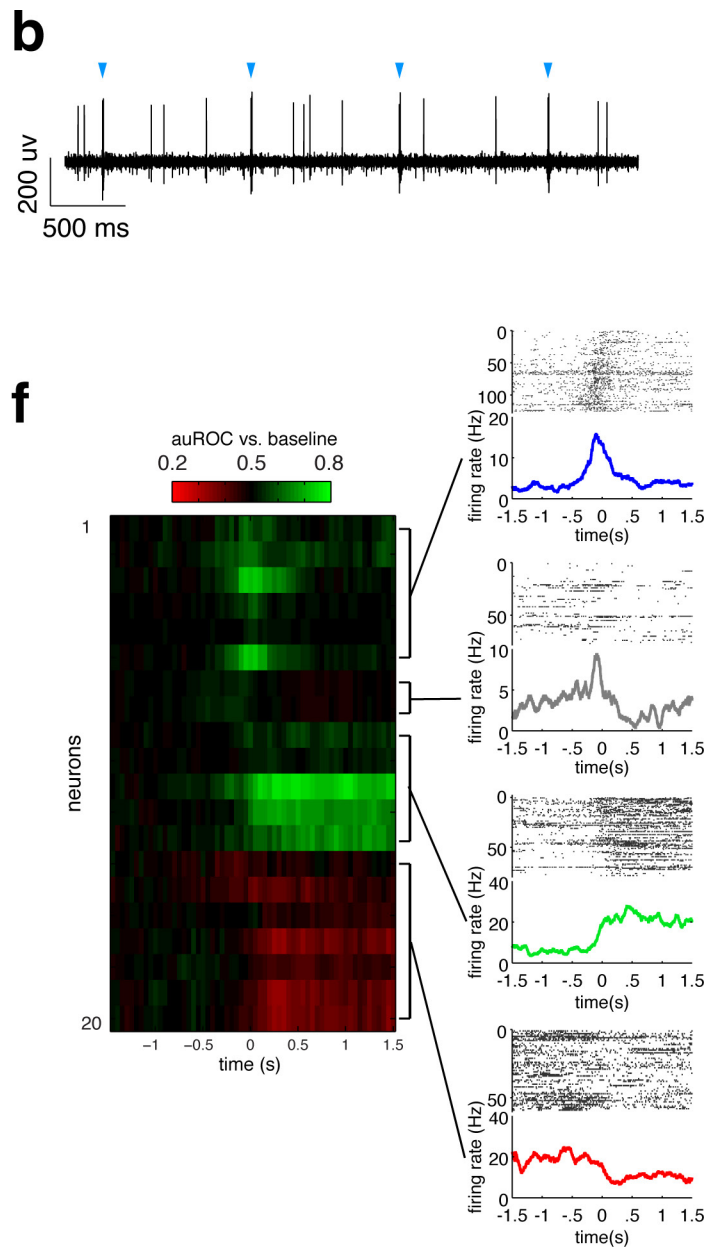
### Extended Data Figure 2 | TH-Cre line and ArchT infection

**characterization.** **a**, TH-Cre mice were crossed with ROSA26R-YFP mice (expression of YFP in Cre<sup>+</sup> cells). This is an example of a midbrain slice of a TH-Cre × ROSA26R-YFP mouse with TH<sup>+</sup> neurons labelled in red. The white line delimits the SNc, and the yellow and green lines delimit areas that cover a depth of 200 μm above and below the nigra, respectively, that were also targeted by stereological cell counts. Scale bar, 100 μm. **b**, Example of a SNc sampling field. Arrowheads denote examples of Cre<sup>+</sup> cells that were TH<sup>+</sup>. Scale bars, 20 μm. **c**, Quantification of the specificity of the Cre line for tagging TH<sup>+</sup> cells ( $n = 3$  slices; 117 counting frames were analysed). **d**, Representative merged image of VTA and SNc after two weeks of infection. ArchT<sup>+</sup> cells are labelled in green, TH<sup>+</sup> cells are labelled in red and merged colours in yellow. ArchT<sup>+</sup> cells are mainly confined to SNc. Scale bar, 500 μm. **e**, Detail of a SNc region labelled for TH (red) and ArchT (green) expression. Arrows are examples

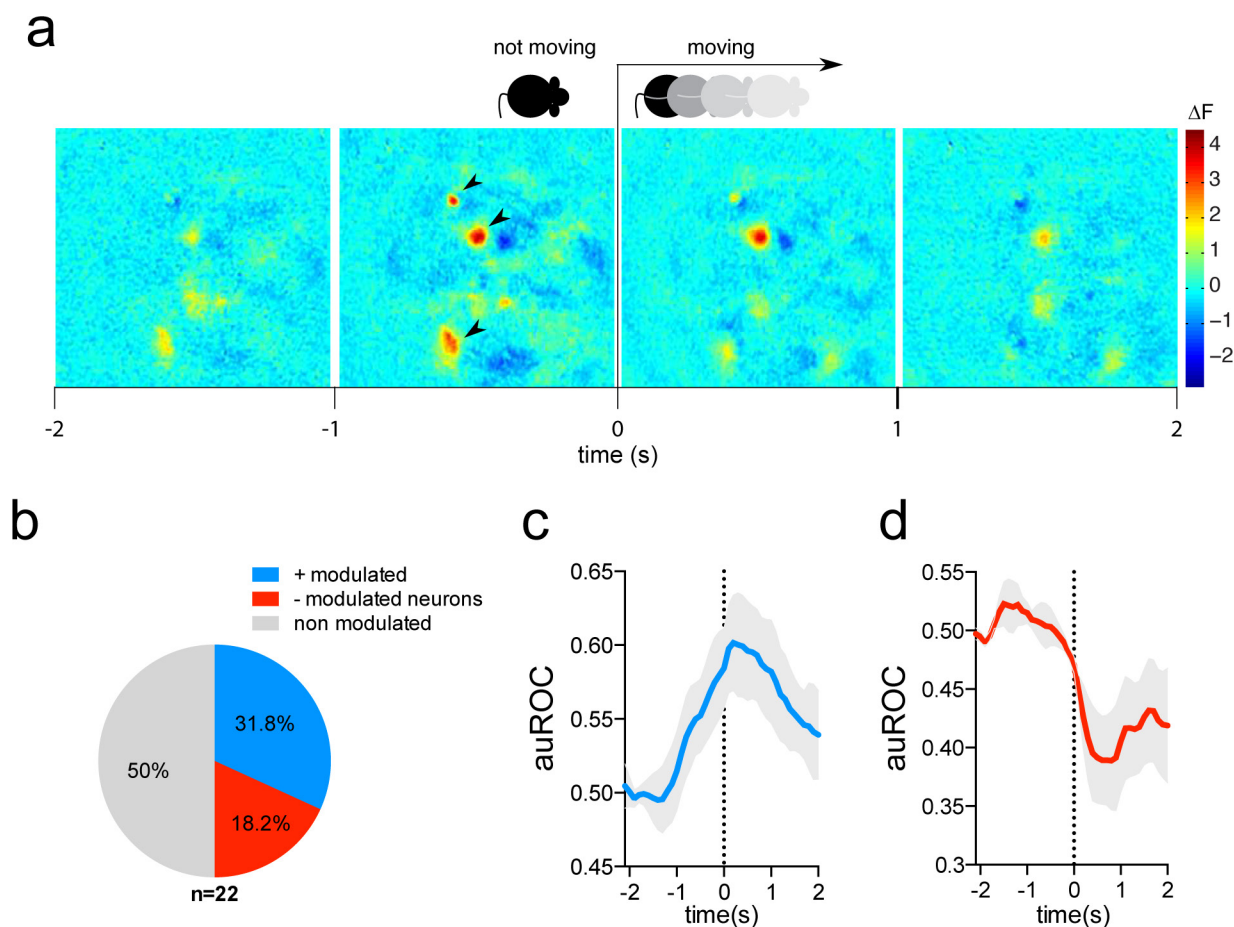
of TH<sup>+</sup> and ArchT<sup>+</sup> cells; closed arrowheads denote examples of TH<sup>+</sup> and ArchT<sup>−</sup> cells. Scale bars, 20 μm. **f**, Efficiency of ArchT virus infection (left). Specificity of ArchT virus infection (right). This was calculated by quantifying the whole SNc stereologically, not only the area closest to the infection ( $n = 6$  slices from two ArchT mice, 122 counting frames). **g**, Examples of the slices and fields used to do the stereological count shown in **f** and **h**. Scale bars, 100 μm. **h**, Stereological quantification of the number of SNc TH<sup>+</sup> cells in YFP- and ArchT-expressing mice after two weeks of infection (ArchT,  $n = 6$  slices from two ArchT mice, 122 counting frames; YFP,  $n = 6$  slices from two YFP mice, 124 counting frames). **i**, Photomicrograph of a midbrain slice of a ArchT-expressing mouse at the end of the experiments (open field and FR8). Red indicates TH<sup>+</sup> cells and green indicates ArchT<sup>+</sup> cells. Scale bar, 100 μm. Data are mean ± s.e.m. (**c**, **f**, **h**).



**Extended Data Figure 3 | Photoidentification and clustering of SNc dopamine neurons.** **a**, Photomicrograph of a midbrain slice of a TH-Cre;Ai32 mouse denoting the right SNc and VTA. ChR2 in green and TH<sup>+</sup> cells in red. Initial electrode position (dashed square) and distance travelled (dashed triangle). Scale bar, 100  $\mu$ m. **b**, Example of continuous recording of a photoidentified neuron. Blue triangles denote 10-ms light pulses of blue light that were delivered at 1 Hz. **c**, PETH of the neuron in **b** aligned to blue-light delivery (100 pulses). **d**, Histogram of latencies to modulation by light delivery. A threshold of 7 ms and an increase in at least 30% firing rate was used to define neurons as photoidentified (blue bars). **e**, Mean spike traces for all photoidentified neurons used in

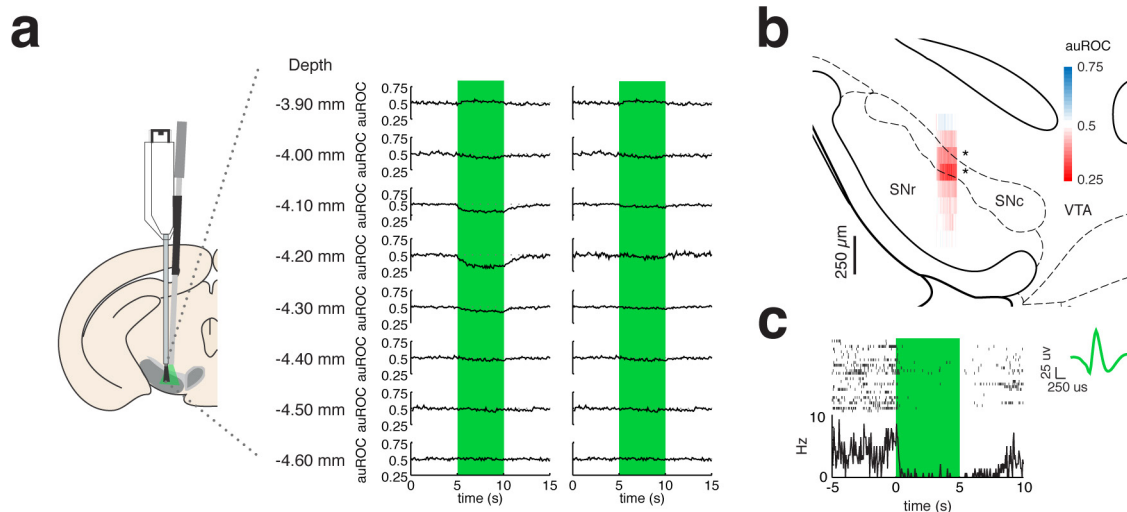


**Fig. 2.** The black trace represents the mean of spikes obtained without light delivery and the blue trace represents the mean trace of spikes obtained during light delivery. **f**, Left, the area under the ROC curve (auROC) was calculated for each time bin of each significantly modulated neuron. Right, we used an affinity propagation algorithm to cluster the traces that resulted from the auROC analysis (see Methods for details). Four clusters were found, of which the PETH of the representative neuron is shown. Neurons were: transiently active before the initiation of movement (blue), transiently active before the initiation of movement followed by inhibition after the initiation (grey), sustained increase in activity with movement initiation (green) or negatively modulated (red).



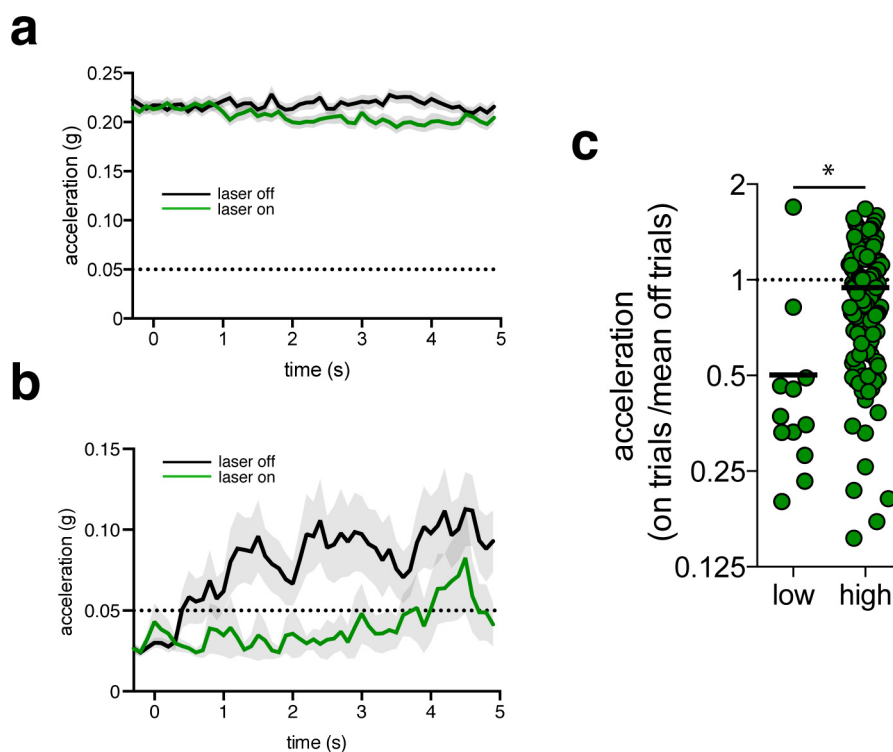
**Extended Data Figure 4 | Calcium imaging of TH<sup>+</sup> neurons from the SNc during movement initiation.** **a**, Example of the average pixel per pixel  $\Delta F$  of one mouse aligned to movement initiation ( $n = 46$  trials). Black arrowheads denote three neurons that are significantly activated before movement initiation. **b**, Proportion of neurons that were positively

or negatively modulated and not modulated by movement initiation ( $n = 22$  neurons obtained from three mice). **c**, Mean auROC trace of positively modulated neurons based on calcium imaging (solid line,  $n = 7$ ). **d**, Mean auROC trace of negatively modulated neurons according to calcium imaging (solid line,  $n = 4$ ). Grey shadow denotes s.e.m. (**c**, **d**).



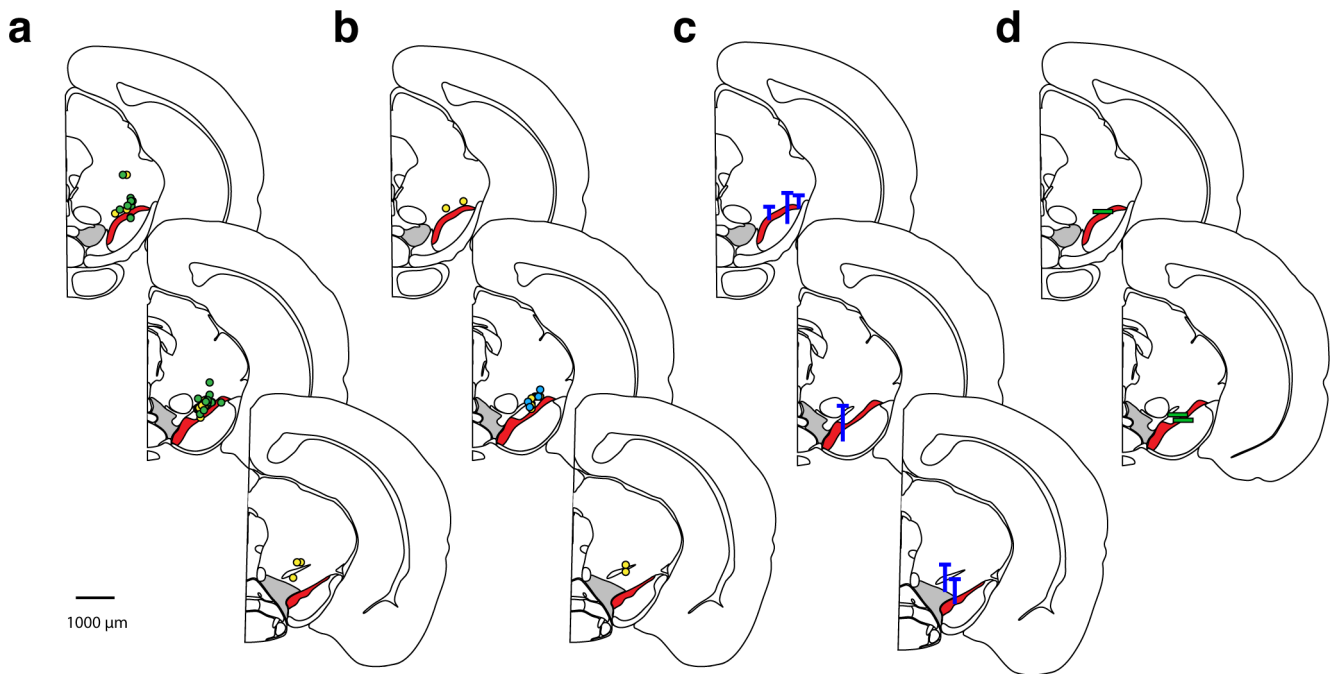
**Extended Data Figure 5 | *In vivo* external recordings reveal specific inhibition of neuronal activity in the SNc.** **a**, Mean unit activity aligned to light onset (values less than 0.5 auROC indicate a decrease compared to baseline and more than 0.5 auROC indicate an increase compared to baseline) at different recording depths. The green rectangle signals the duration of light delivery. Left, mean of all units recorded. Right, mean of all units except negatively modulated units. **b**, Top, anatomical representation<sup>31</sup> of the mean unit activity depending on recording depth and the location of the cannula of the recording electrode (red, decrease from baseline; blue, increase from baseline). The percentage of inhibited cells was not homogeneous throughout all depths ( $\chi^2_{4,140} = 18.01$ ,  $P < 0.05$ , test based on five levels of depth from -3.9 to -4.6 mm with 150- $\mu$ m steps). In fact, when we investigated the mean activity of all units recorded at

each depth, we found that the mean activity during light delivery changed depending on the depth, and it was only significantly different at the depths at which the SNc is located, for which the percentage of inhibited units was 61.3%. This is anatomically represented in **b**. Depth (number of neurons): -3.9 mm (3); -4 mm (22); -4.1 mm (14); -4.2 mm (6); -4.3 mm (8); -4.4 mm (10); -4.5 mm (4); -4.6 mm (5). Kruskal-Wallis test:  $H = 18.22$ ;  $P = 0.011$ . Dunn's multiple comparison test, all means compared to mean at -4.6 mm: 3.9 mm,  $P > 0.99$ ; -4 mm,  $P = 0.078$ ; -4.1 mm,  $*P = 0.017$ ; -4.2 mm,  $**P = 0.008$ ; -4.3 mm,  $P = 0.67$ ; -4.4 mm,  $P = 0.82$ ; -4.5 mm,  $P > 0.99$ ). Asterisks indicate depths with mean auROC significantly different from -4.6 mm depth. **c**, Example of a single unit inhibited by green light. The mouse brain has been reproduced with permission from ref. 31.



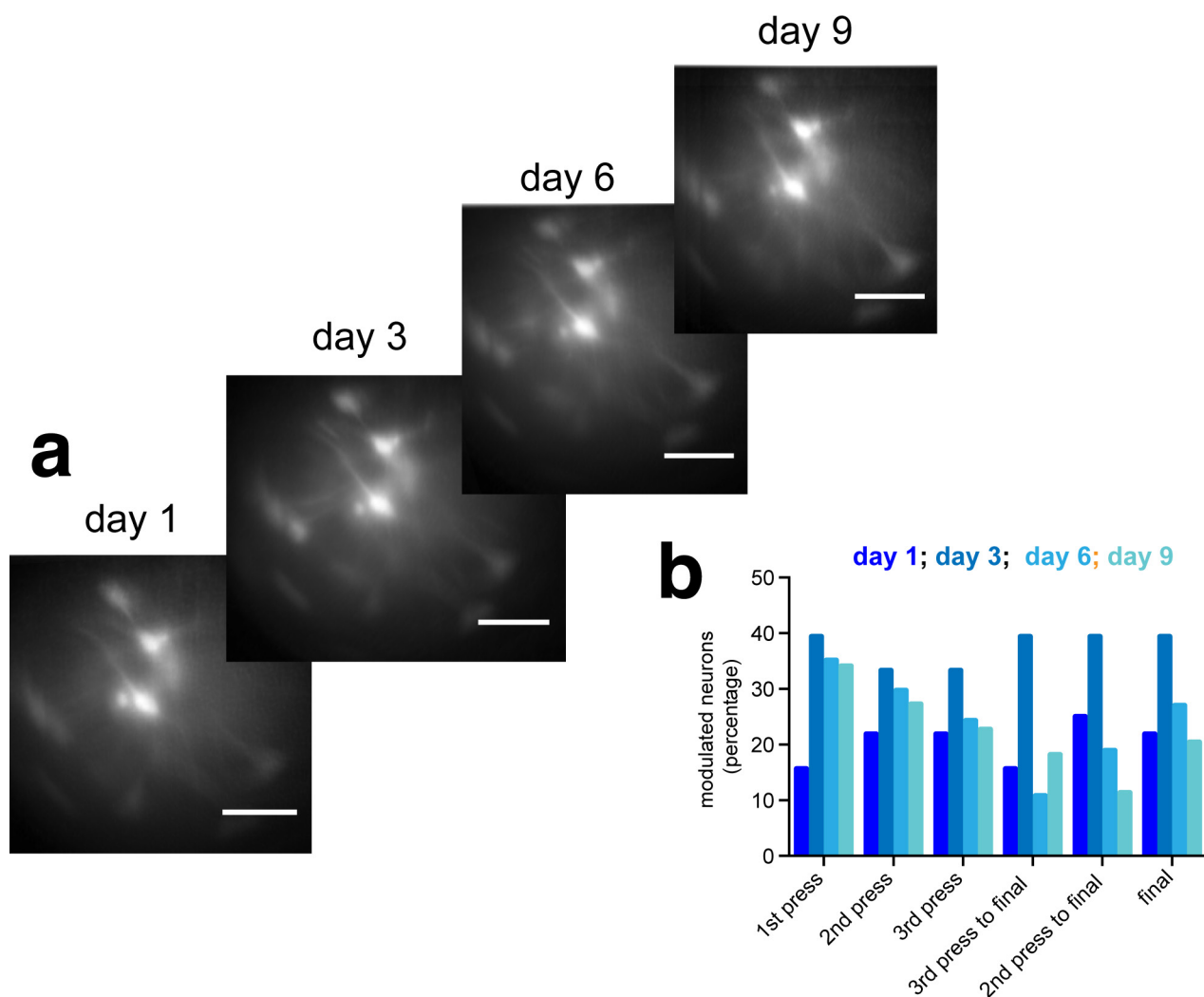
**Extended Data Figure 6 | Five-second inhibition of SNc TH<sup>+</sup> neurons during mobile and immobile trials.** **a**, Acceleration during laser-off and brief laser-on trials (5 s inhibition) when ArchT mice were mobile before the start of the trial ( $n = 217$  laser-on trials and  $n = 212$  laser-off trials obtained from 11 ArchT mice). The horizontal dotted line denotes the threshold used to classify acceleration state. **b**, Acceleration during laser-off and brief laser-on trials (5 s inhibition) when ArchT mice were immobile before trial start ( $n = 17$  laser-on trials from 5 ArchT mice and  $n = 12$  laser-off trials obtained from 7 ArchT mice). Horizontal dotted line

denotes the threshold used to classify acceleration state. **c**, Acceleration during brief laser-on (5 s inhibition) normalized to mean laser-off acceleration for immobile and mobile states ( $n = 17$  laser-on immobile trials obtained from 5 ArchT mice,  $n = 12$  laser-off immobile trials obtained from 7 ArchT mice;  $n = 217$  laser-on mobile trials,  $n = 212$  laser-off mobile trials obtained from 11 mice). Acceleration state significantly affected normalized acceleration (linear mixed model with 'mouse' as a random effect,  $F = 19.57$ ,  $P < 0.0001$ ).



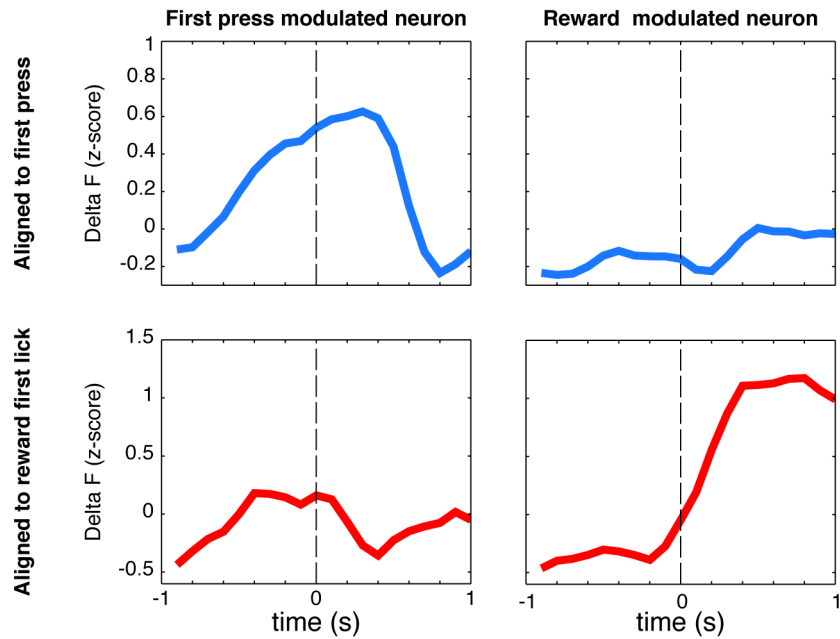
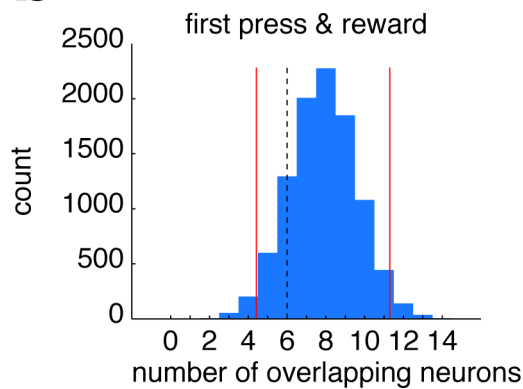
**Extended Data Figure 7 | Anatomical position of fibres, electrodes and lens.** **a**, Fibre placement (green, ArchT mice; yellow, YFP mice). **b**, Fibre placement (blue, ChR2 mice; yellow, YFP mice). **c**, Optrode placement. Horizontal blue lines denotes the cannula position and vertical lines the distance travelled by the electrodes. **d**, GRIN lens placement (green bar indicates the position of the bottom of the lens). The representations of

the mouse brain and anatomical structures was obtained from the Allen Mouse Brain Atlas (2004) using API. Top, [http://api.brain-map.org/api/v2/svg\\_download/100960073?groups=28](http://api.brain-map.org/api/v2/svg_download/100960073?groups=28); middle, [http://api.brain-map.org/api/v2/svg\\_download/100960057?groups=28](http://api.brain-map.org/api/v2/svg_download/100960057?groups=28); bottom, [http://api.brain-map.org/api/v2/svg\\_download/100960525?groups=28](http://api.brain-map.org/api/v2/svg_download/100960525?groups=28).



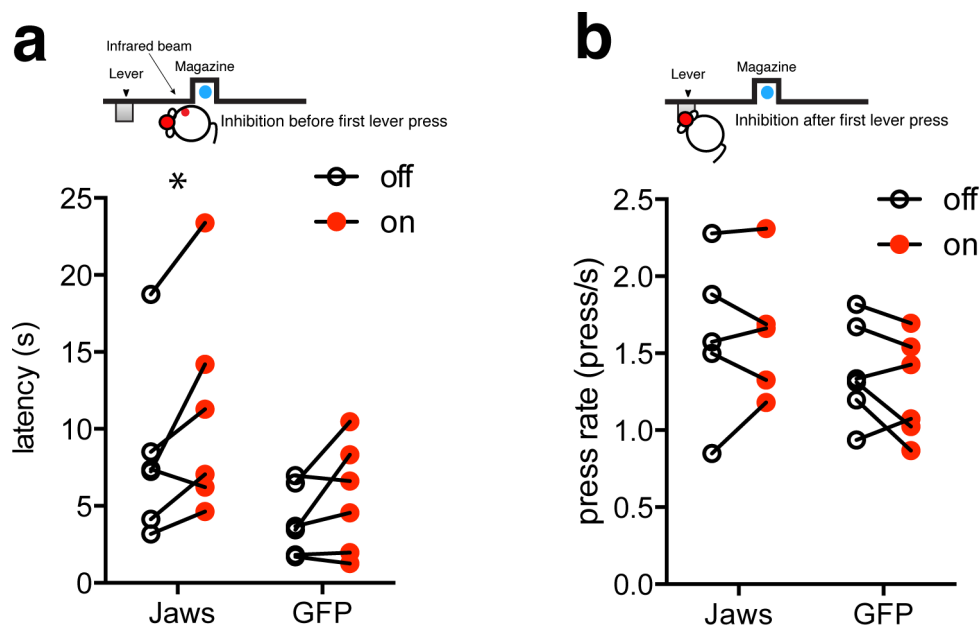
**Extended Data Figure 8 | Lever-press significant neurons across training days. a,** Field-of view of one of the GCaMP6-expressing mice across different training days spanning days from the beginning to the end of training. Scale bars, 100  $\mu$ m. **b,** Percentage of neurons significantly

modulated by the sequence lever presses throughout training. Data shown were obtained during the training of three out of the four mice used to obtain the calcium imaging data shown in Fig. 4. Data during training was not available for one mouse.

**a****b**

**Extended Data Figure 9 | First press- and reward-related neurons populations do not overlap.** **a**, Example of a neuron modulated by first press but not reward (left) and a neuron related to reward but not first press (right), aligned to first press (blue) and reward consumption (red). **b**, Monte Carlo simulations (10,000 samples) were used to generate

a distribution of the number of overlapping neurons for first press and reward, assuming random assignment. Red lines denote the 95% confidence interval. Dashed line represents the number of overlapping neurons found in our experiment.



**Extended Data Figure 10 | FR8-inhibition experiment using Jaws.** We replicated the result obtained in the FR8 task (Fig. 4h–j), using Jaws (see Methods for details). **a**, Latency to initiate lever press sequence for laser-off trials and trials with inhibition starting just before sequence initiation for both Jaws ( $n = 6$ ) and GFP ( $n = 6$ ) groups. Two-way mixed ANOVA; planned comparisons between laser-on and laser-off trials using Fisher's least significant difference tests; main effect group  $F_{1,10} = 3.16$ ,  $P = 0.11$ ; main effect laser  $F_{1,10} = 9.074$ ,  $P = 0.0131$ ; interaction effect  $F_{1,10} = 0.92$ ,  $P = 0.36$ ; planned comparisons: Jaws laser off – laser on

$P = 0.019$ , GFP laser off – laser on  $P = 0.18$ . **b**, Press rate in trials with no light delivery and trials with light delivery starting after the first press for both Jaws ( $n = 5$ ) and GFP ( $n = 6$ ) groups. Two-way mixed ANOVA; planned comparisons between laser-on and laser-off trials using Fisher's least significant difference tests; main effect group  $F_{1,9} = 1.607$ ,  $P = 0.24$ ; main effect laser  $F_{1,9} = 0.53$ ,  $P = 0.49$ ; interaction effect  $F_{1,9} = 1.01$ ,  $P = 0.34$ ; planned comparisons: Jaws laser off – laser on  $P = 0.86$ , GFP laser off – laser on  $P = 0.23$ .

# High performance plasma amyloid- $\beta$ biomarkers for Alzheimer's disease

Akinori Nakamura<sup>1</sup>, Naoki Kaneko<sup>2</sup>, Victor L. Villemagne<sup>3,4</sup>, Takashi Kato<sup>1,5</sup>, James Doecke<sup>6</sup>, Vincent Doré<sup>3,6</sup>, Chris Fowler<sup>4</sup>, Qiao-Xin Li<sup>4</sup>, Ralph Martins<sup>7</sup>, Christopher Rowe<sup>3,4</sup>, Taisuke Tomita<sup>8</sup>, Katsumi Matsuzaki<sup>9</sup>, Kenji Ishii<sup>10</sup>, Kazunari Ishii<sup>11</sup>, Yutaka Arahata<sup>5</sup>, Shinichi Iwamoto<sup>2</sup>, Kengo Ito<sup>1,5</sup>, Koichi Tanaka<sup>2</sup>, Colin L. Masters<sup>4</sup> & Katsuhiko Yanagisawa<sup>1</sup>

**To facilitate clinical trials of disease-modifying therapies for Alzheimer's disease, which are expected to be most efficacious at the earliest and mildest stages of the disease<sup>1,2</sup>, supportive biomarker information is necessary. The only validated methods for identifying amyloid- $\beta$  deposition in the brain—the earliest pathological signature of Alzheimer's disease—are amyloid- $\beta$  positron-emission tomography (PET) imaging or measurement of amyloid- $\beta$  in cerebrospinal fluid. Therefore, a minimally invasive, cost-effective blood-based biomarker is desirable<sup>3,4</sup>. Despite much effort<sup>3–7</sup>, to our knowledge, no study has validated the clinical utility of blood-based amyloid- $\beta$  markers. Here we demonstrate the measurement of high-performance plasma amyloid- $\beta$  biomarkers by immunoprecipitation coupled with mass spectrometry. The ability of amyloid- $\beta$  precursor protein (APP)<sub>669–711</sub>/amyloid- $\beta$  (A $\beta$ )<sub>1–42</sub> and A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> ratios, and their composites, to predict individual brain amyloid- $\beta$ -positive or -negative status was determined by amyloid- $\beta$ -PET imaging and tested using two independent data sets: a discovery data set (Japan,  $n = 121$ ) and a validation data set (Australia,  $n = 252$  including 111 individuals diagnosed using <sup>11</sup>C-labelled Pittsburgh compound-B (PIB)-PET and 141 using other ligands). Both data sets included cognitively normal individuals, individuals with mild cognitive impairment and individuals with Alzheimer's disease. All test biomarkers showed high performance when predicting brain amyloid- $\beta$  burden. In particular, the composite biomarker showed very high areas under the receiver operating characteristic curves (AUCs) in both data sets (discovery, 96.7%,  $n = 121$  and validation, 94.1%,  $n = 111$ ) with an accuracy approximately equal to 90% when using PIB-PET as a standard of truth. Furthermore, test biomarkers were correlated with amyloid- $\beta$ -PET burden and levels of A $\beta$ <sub>1–42</sub> in cerebrospinal fluid. These results demonstrate the potential clinical utility of plasma biomarkers in predicting brain amyloid- $\beta$  burden at an individual level. These plasma biomarkers also have cost-benefit and scalability advantages over current techniques, potentially enabling broader clinical access and efficient population screening.**

Attempts to use conventional enzyme-linked immunosorbent assay (ELISA)-based techniques to assess plasma amyloid- $\beta$  (A $\beta$ ) levels in patients have not been successful (see Supplementary Information for more detailed background information). Immunoprecipitation–mass spectrometry (IP–MS) assays have been proposed<sup>8,9</sup> as an alternative, although the sample sizes in both of these studies were small ( $n = 62$  and  $n = 41$ , respectively). Using IP–MS, we originally developed a plasma biomarker that discriminated individuals with high levels of A $\beta$  (A $\beta$ <sup>+</sup>) from individuals with low levels (A $\beta$ <sup>−</sup>) with more than 90% sensitivity and specificity when classified using PIB-PET<sup>8</sup>. In that study,

we used IP–MS with matrix-assisted laser desorption ionization–time-of-flight (MALDI–TOF) mass spectrometry, which can also be used for protein quantification<sup>10,11</sup>, to measure the ratio of plasma A $\beta$ <sub>1–42</sub> to a novel APP<sub>669–711</sub> fragment (APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub>) (Extended Data Fig. 1a). Here we improved the general applicability and reproducibility of the previous IP–MS methodology through exploratory studies. We found that the ratio of A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> also performed at the same level as APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub>, and that a composite biomarker score that incorporated both could further improve performance (Supplementary Information and Extended Data Fig. 1b). Thus, we hypothesized that APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub>, A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> and the composite biomarker generated by the IP–MS assay were promising and potentially clinically useful candidates for plasma biomarkers as surrogates for brain A $\beta$  burden. Our retrospective cross-sectional study tested this hypothesis in a discovery data set from the Japanese National Center for Geriatrics and Gerontology (NCGG) (121 samples), and was externally validated using an independent data set derived from the Australian Imaging, Biomarker and Lifestyle Study of Ageing (AIBL)<sup>12</sup> cohort (252 samples) (Table 1). Both data sets include a balanced number of individuals clinically classified as cognitively normal, individuals with mild cognitive impairment (MCI) and individuals clinically diagnosed with Alzheimer's disease (AD) with dementia. All samples had corresponding A $\beta$ -PET data obtained using PIB (NCGG and AIBL), flutemetamol (FLUTE) or florbetapir (FBP) (AIBL). Information on the levels of A $\beta$  in cerebrospinal fluid (CSF A $\beta$ ) was available for a subset of the AIBL cohort. The primary aim of the study was to assess the performance of plasma-A $\beta$  biomarkers for determining an individual's status of A $\beta$  deposition, using PIB-PET as the standard of truth. For secondary outcomes, we examined the performance of the plasma-A $\beta$  biomarker against other PET ligands (FLUTE and FBP) and within clinical categories (cognitively normal, MCI, and AD). We also examined the correlations of plasma-A $\beta$  biomarkers with A $\beta$ -PET burden and CSF A $\beta$  values.

Figure 1 and Extended Data Fig. 2a show the normalized intensity of plasma A $\beta$  as measured by IP–MS and the values of test biomarkers for each study site. The test biomarker values were generated by computing the ratio of the normalized intensity of the peptides. A $\beta$ <sub>1–42</sub> was used as denominator (APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub> and A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub>), because it yielded normal distributions (Extended Data Fig. 2b). The composite biomarker was generated by combining normalized scores of APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub> and A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> with a pre-determined weight of 1:1 (Methods, Supplementary Information and Extended Data Fig. 1b). All of the test biomarkers showed highly significant differences ( $P < 0.0001$ , two-sided Student's  $t$ -test or Welch's  $t$ -test) between the A $\beta$ <sup>+</sup> and A $\beta$ <sup>−</sup> groups (Extended Data Fig. 2a). At a single peptide level,

<sup>1</sup>Center for Development of Advanced Medicine for Dementia, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. <sup>2</sup>Koichi Tanaka Mass Spectrometry Research Laboratory, Shimadzu Corporation, Kyoto 604-8511, Japan. <sup>3</sup>Austin Health, Department of Molecular Imaging and Therapy, Center for PET, Heidelberg, Victoria 3084, Australia. <sup>4</sup>The Florey Institute, The University of Melbourne, Parkville 3010, Australia. <sup>5</sup>National Hospital for Geriatric Medicine, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. <sup>6</sup>Health and Biosecurity, CSIRO, Brisbane 4029, Australia. <sup>7</sup>Edith Cowan University, Joondalup, Western Australia 6027, Australia. <sup>8</sup>Laboratory of Neuropathology and Neuroscience, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033, Japan. <sup>9</sup>Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan. <sup>10</sup>Team for Neuroimaging Research, Tokyo Metropolitan Institute of Gerontology, Tokyo 173-0015, Japan. <sup>11</sup>Department of Radiology, Kindai University Faculty of Medicine, Osakasayama, Osaka 589-8511, Japan.

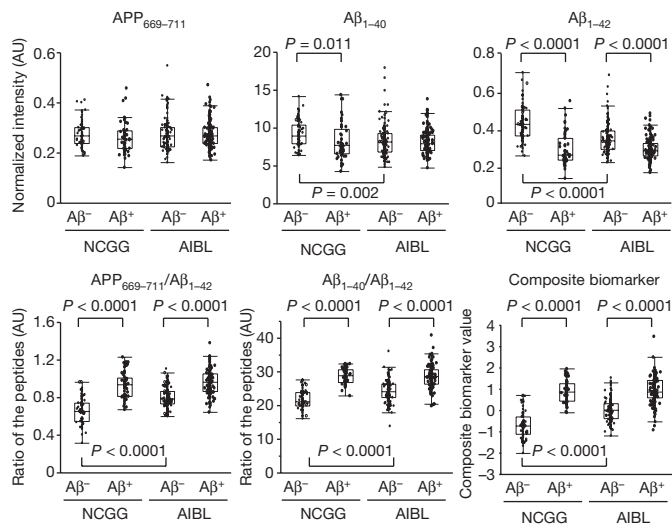
**Table 1 | Demographics of the subjects in each study site (NCGG and AIBL)**

|  | NCGG            |                     |                 | AIBL            |                 |                 |
|--|-----------------|---------------------|-----------------|-----------------|-----------------|-----------------|
| PET tracer   | PIB             | PIB                 | FLUTE           | FBP             | AIBL overall    | CSF             |
| Sample size (n) total  | 121             | 111                 | 81              | 60              | 252             | 46              |
| A $\beta$ <sup>+</sup> /A $\beta$ <sup>-</sup>                           | 50/71           | 60/51               | 47/34           | 30/30           | 137/115         | 25/21           |
| CN/MCI/AD (A $\beta$ <sup>+</sup> + A $\beta$ <sup>-</sup> )             | 62/30/29        | 63/33/15            | 43/30/8         | 50/4/6          | 156/67/29       | 30/9/7          |
| CN/MCI/AD (A $\beta$ <sup>+</sup> )                                      | 10/20/20        | 25/20/15*           | 20/19/8         | 21/3/6          | 66/42/29        | 13/5/7          |
| CN/MCI/AD (A $\beta$ <sup>-</sup> )                                      | 52/10/9         | 38/13/0**           | 23/11/0         | 29/1/0          | 90/25/0         | 17/4/0          |
| Age (A $\beta$ <sup>+</sup> + A $\beta$ <sup>-</sup> , mean $\pm$ s.d.)  | 74.0 $\pm$ 5.1  | 75.3 $\pm$ 6.5      | 72.1 $\pm$ 4.5  | 74.8 $\pm$ 5.2  | 74.2 $\pm$ 5.8  | 73.7 $\pm$ 5.5  |
| Age (A $\beta$ <sup>+</sup> , mean $\pm$ s.d.)                           | 75.3 $\pm$ 4.7  | 75.3 $\pm$ 6.3      | 72.1 $\pm$ 4.4  | 75.7 $\pm$ 4.8  | 74.3 $\pm$ 5.6  | 72.4 $\pm$ 4.2  |
| Age (A $\beta$ <sup>-</sup> , mean $\pm$ s.d.)                           | 73.0 $\pm$ 5.2  | 75.4 $\pm$ 6.8      | 72.0 $\pm$ 4.5  | 73.9 $\pm$ 5.4  | 74.0 $\pm$ 6.0  | 75.1 $\pm$ 6.4  |
| Gender (A $\beta$ <sup>+</sup> + A $\beta$ <sup>-</sup> , M/F)           | 55/66           | 56/55               | 40/41           | 33/27           | 129/123         | 26/20           |
| Gender (A $\beta$ <sup>+</sup> , M/F)                                    | 22/28           | 32/28               | 28/19           | 11/19           | 71/66           | 18/7            |
| Gender (A $\beta$ <sup>-</sup> , M/F)                                    | 31/40           | 24/27               | 12/22           | 22/8            | 58/57           | 11/10           |
| APOE4 (A $\beta$ <sup>+</sup> + A $\beta$ <sup>-</sup> , +/–)            | 50/71           | 53/58               | 34/47           | 21/39           | 108/144         | 15/31           |
| APOE4 (A $\beta$ <sup>+</sup> , +/–)                                     | 35/15           | 42/18               | 30/17           | 15/15           | 87/50           | 9/16            |
| APOE4 (A $\beta$ <sup>-</sup> , +/–)                                     | 15/56           | 11/40               | 4/30            | 6/24            | 21/94           | 2/19            |
| SUVr (A $\beta$ <sup>+</sup> + A $\beta$ <sup>-</sup> , mean $\pm$ s.d.) | 1.52 $\pm$ 0.51 | 1.75 $\pm$ 0.61***  | 0.71 $\pm$ 0.23 | 1.12 $\pm$ 0.21 | 1.71 $\pm$ 0.55 | 1.64 $\pm$ 0.47 |
| SUVr (A $\beta$ <sup>+</sup> , mean $\pm$ s.d.)                          | 2.05 $\pm$ 0.37 | 2.24 $\pm$ 0.37**** | 0.86 $\pm$ 0.2  | 1.29 $\pm$ 0.17 | 2.11 $\pm$ 0.43 | 1.98 $\pm$ 0.37 |
| SUVr (A $\beta$ <sup>-</sup> , mean $\pm$ s.d.)                          | 1.14 $\pm$ 0.10 | 1.17 $\pm$ 0.08     | 0.51 $\pm$ 0.02 | 0.95 $\pm$ 0.05 | 1.23 $\pm$ 0.09 | 1.24 $\pm$ 0.10 |

Breakdown of the number of subjects for each variable, with the exception of age and standardized uptake value ratio (SUVr) values. SUVr values represent SUVr for PIB, FLUTE and FBP, and SUVr/BeCKeT (before the centiloid kernel transformation<sup>17</sup>) values for AIBL overall and CSF. Site differences between NCGG and AIBL were tested only for PIB-PET groups using Student's *t*-test (age and SUVr) or  $\chi^2$  test (all others). The CSF group is a subset of the AIBL data including 17 PIB, 18 FLUTE, and 11 FBP cases. Asterisks indicate statistically significant site differences: \**P* = 0.043, \*\**P* = 0.014, \*\*\**P* = 0.002, \*\*\*\**P* = 0.007; two-sided tests.

A $\beta$ <sub>1–42</sub> also showed highly significant group differences (*P* < 0.0001), whereas APP<sub>669–711</sub> did not show any group differences, and A $\beta$ <sub>1–40</sub> showed a group difference in the NCGG data set (*P* = 0.011), but not in the AIBL data set. Significant (*P* < 0.05) site differences between the NCGG and AIBL data sets were seen for all peptides and biomarkers except for APP<sub>669–711</sub>.

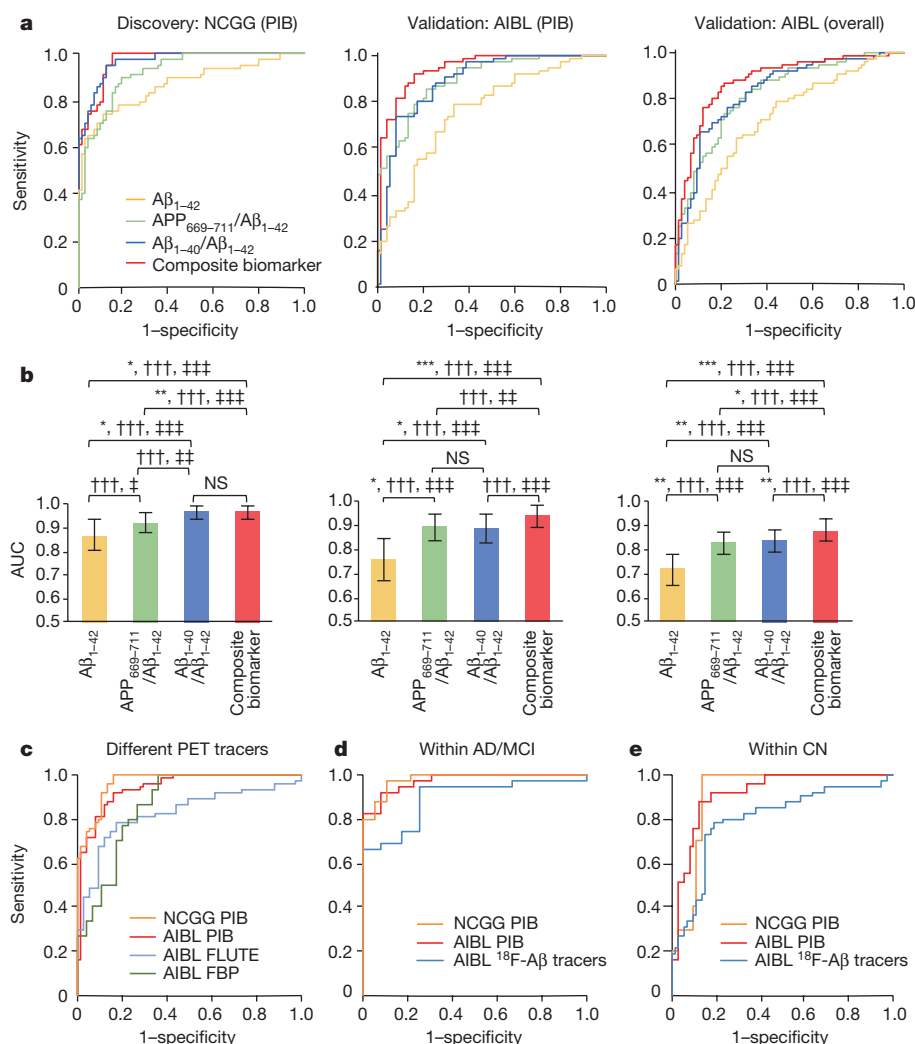
To evaluate the performance of plasma biomarkers in predicting brain A $\beta$  burden, we conducted receiver operating characteristic (ROC) analyses with the discovery and validation data sets (Fig. 2a and Extended Data Table 1a, left). A $\beta$ <sub>1–42</sub> peptide alone showed moderately high areas under the curves (AUCs) in the discovery (NCGG) and validation (AIBL PIB and AIBL overall) analyses with values (87.2%, 75.7% and 71.8% for NCGG, AIBL PIB and AIBL overall, respectively) far beyond the chance level of AUC = 50% (asymptotic significance, *P* < 0.0001). Compared with A $\beta$ <sub>1–42</sub>, all of the test biomarkers (APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub>, A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> and the composite biomarker) showed significantly better predictive ability as evaluated by the net reclassification



**Figure 1 | The peptide and biomarker values in each study site.** Box plots showing each peptide (upper), and test biomarker (lower) value in the NCGG (*n* = 121) and AIBL overall (*n* = 252) data sets. Significant group differences are indicated by *P* values (two-sided Student's *t*-test or Welch's *t*-test). The boxes represent the 25th, 50th (median) and 75th percentiles of the data; the whiskers represent the lowest (or highest) datum within 1.5  $\times$  interquartile range from the 25th (or 75th) percentile. See Extended Data Fig. 2a for detailed values. AU, arbitrary units.

improvement (NRI) and the integrated discrimination improvement (IDI) (see Methods) in all analyses (Bonferroni-corrected *P* < 0.05) (Fig. 2b). In addition, the AUCs of these three test biomarkers were significantly higher than those of A $\beta$ <sub>1–42</sub> in all analyses (DeLong test, Bonferroni-corrected *P* < 0.05) (see Methods) except for APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub> in the NCGG data set. The composite biomarker showed the highest AUCs in all analyses (96.7%, 94.1% and 88.3%, respectively, for NCGG, AIBL PIB and AIBL overall). In the AIBL PIB and overall analyses, the composite biomarker showed significant improvements in NRI and IDI compared with both APP<sub>669–711</sub>/A $\beta$ <sub>1–42</sub> and A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> (Bonferroni-corrected *P* < 0.01) (Fig. 2b). In the NCGG data set, A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> showed identically high performance to the composite biomarker. Comparisons between the NCGG PIB and AIBL PIB analyses demonstrated that performances were generally lower in the validation analyses, especially for A $\beta$ <sub>1–40</sub>/A $\beta$ <sub>1–42</sub> (DeLong test, uncorrected *P* = 0.026); however, the composite biomarker showed similarly high performances with an AUC of approximately 95% and approximately 90% accuracy. The AIBL overall (all PET tracers) analyses showed slightly lower performances compared with the AIBL PIB analyses. The biomarker performances in the analyses adjusted for age, gender, clinical category and the presence of the APOE- $\epsilon$ 4 (APOE4) allele showed a similar tendency to the unadjusted analyses, while the adjusted analyses generally showed slightly higher AUCs than the unadjusted analyses (Extended Data Fig. 3a, b and Extended Data Table 1a, right).

As the composite biomarker showed the highest and most stable performance across all analyses, it was the main focus of subsequent tests. We further analysed the performance of the composite biomarker against different A $\beta$ -PET tracers. When the <sup>18</sup>F-A $\beta$  ligands FLUTE and FBP were used to classify participants into A $\beta$ <sup>+</sup> or A $\beta$ <sup>-</sup> groups, the performances of the biomarkers were slightly lower than those obtained with PIB (Fig. 2c and Extended Data Table 1b, left). Within the AIBL data set, the AUCs of the composite biomarker for FLUTE (82.9%) and FBP (86.4%) were lower than for PIB (94.1%) in the unadjusted analyses (DeLong test, uncorrected *P* = 0.033 and 0.149 for FLUTE and FBP, respectively). The adjusted (age, gender, APOE4, and clinical category) analyses showed similar results (Extended Data Fig. 3c and Extended Data Table 1b, right). Given that there were no significant differences between the two independent PIB data sets, we consider it to be unlikely that variability in the biomarker performance causes the lower relative performance observed with <sup>18</sup>F-A $\beta$  tracers. It may instead be the consequence of the higher variance and lower performance of the <sup>18</sup>F-A $\beta$  tracers compared to PIB<sup>13–16</sup> (see Supplementary Discussion).



**Figure 2 | High performance of the plasma biomarkers.** **a**, ROC analyses for each biomarker when predicting individual  $A\beta^+/A\beta^-$  status for the discovery and validation data sets. Unadjusted analyses of the NCGG PIB discovery data (left), the AIBL PIB (middle) and AIBL overall (all tracers, right) validation data. See Extended Data Table 1a for detailed performance values. Data are from 121, 111 and 252 individuals for the NCGG PIB, AIBL and AIBL overall data, respectively. **b**, Comparisons of biomarker performances within each analysis corresponding to the ROC curves in **a**. Each colour bar represents the AUC and 95% confidence interval. Statistically significant differences between two AUCs (DeLong test) are indicated by asterisks: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Significant increments in predictive ability as assessed by NRI and IDI are indicated by daggers and double daggers, respectively. † or ‡ $P < 0.05$ ; †† or ††† $P < 0.01$ ; ††† or †††† $P < 0.001$ . All  $P$  values are two-sided and

Bonferroni corrected (multiplied by the number of comparisons, 6).

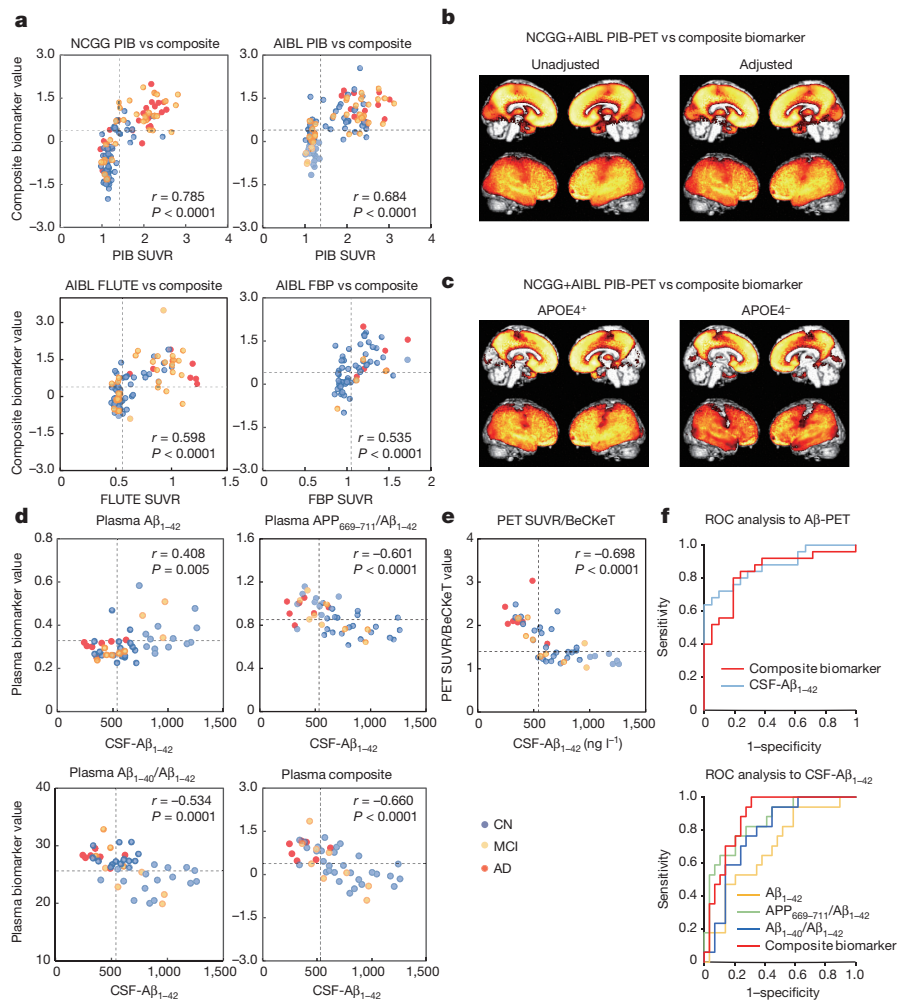
NS, not significant. **c**, Unadjusted ROC analyses of the composite biomarker compared with different PET tracers; PIB (NCGG,  $n = 121$ , and AIBL,  $n = 111$ ), flutemetamol (AIBL,  $n = 81$ ), and florbetapir (AIBL,  $n = 60$ ). See Extended Data Table 1b for detailed performance values.

**d**, **e**, Unadjusted ROC curves of the composite biomarker within the AD and MCI (**d**), and cognitively normal (**e**) groups. For the AD and MCI group, data are from 59 individuals for NCGG PIB and from 48 individuals for AIBL PIB and the  $^{18}F$ -A $\beta$  tracers. For the analyses of the cognitively normal group, data are from 62, 63 and 93 individuals for NCGG PIB, AIBL PIB and the  $^{18}F$  A $\beta$  tracers, respectively. See Extended Data Table 1c for detailed performance values; corresponding results of the adjusted analyses for **a–e** are shown in Extended Data Fig. 3.

Plasma-biomarker performances were also evaluated by clinical category. To obtain a sufficient number in each subpopulation (as estimated by power analysis; see Methods), AD and MCI were analysed as one group (AD/MCI), and compared with the cognitively normal group. The data for FLUTE and FBP were also grouped and analysed together as  $^{18}F$ -A $\beta$  tracers. The results of the ROC analyses for the composite biomarker are shown in Fig. 2d, e and Extended Data Table 1c (left). Within the clinical category of AD/MCI, the performance of the composite biomarker against PIB and  $^{18}F$ -A $\beta$  tracers was very high, with AUCs of 97.4% and 89.4% and accuracies of 91.7% and 89.6%, respectively, in the unadjusted analysis for the validation AIBL data. Within the cognitively normal group, performance against PIB was also high (AUC = 91.7%, accuracy = 87.3%); however, performance against  $^{18}F$ -A $\beta$  tracers was considerably lower (AUC = 80.0%, accuracy = 79.6%), and it did not

reach significance (DeLong test, uncorrected  $P = 0.053$ ). Adjusted (age, gender and  $APOE4$ ) analyses showed similar results (Extended Data Fig. 3d, e and Extended Data Table 1c, right).

To evaluate the strength of the link between plasma biomarkers and A $\beta$ -PET burden, we conducted correlation analyses. All of the plasma biomarkers, including  $A\beta_{1-42}$  peptide alone, showed significant correlations with A $\beta$ -PET burden (Fig. 3a and Extended Data Fig. 4a–d). The strongest correlations were found between PIB standardized uptake value ratio (SUVR) and the composite biomarker in the NCGG, AIBL and NCGG + AIBL combined data sets, with correlation coefficients of  $r = 0.785$ ,  $0.684$  and  $0.735$ , respectively (Pearson's correlation coefficient, all  $P < 0.0001$ ). The correlation coefficients against FLUTE-SUVR ( $r = 0.598$ ,  $P < 0.0001$ ) and FBP-SUVR ( $r = 0.535$ ,  $P < 0.0001$ ) were slightly lower than those observed with PIB. The correlation



**Figure 3 | Plasma biomarkers are significantly correlated with brain Aβ burden and CSF-Aβ<sub>1-42</sub> level.** **a**, Composite biomarker values plotted against SUVR values of Aβ-PET imaging for each tracer. Data are from 121, 111, 81 and 60 individuals for the NCGG PiB, AIBL PiB, AIBL FLUTE and AIBL FBP analyses, respectively (see also Extended Data Fig. 4). Note that in the NCGG data, there are nine patients with AD who were clinically diagnosed as having AD but were PIB-PET classified as Aβ<sup>-</sup> and thus probably did not have AD. **b**, Topographical associations between the composite biomarker values and cerebral Aβ burden as assessed by Aβ-PET imaging. Unadjusted (left) and adjusted (age and APOE4, right) regression analyses were performed with the NCGG + AIBL PIB-PET data set ( $n = 232$ ). **c**, Regression analyses between Aβ-PET and the composite biomarker within APOE4 positive ( $n = 103$ ) and negative ( $n = 129$ ) individuals from the NCGG + AIBL PIB-PET data set. For **b** and **c**, brain regions that showed significantly positive correlations (FWE corrected  $P < 0.05$ ) are visualized. **d**, Scatter plots for the CSF Aβ<sub>1-42</sub> level ( $n = 46$ ) and plasma biomarker values. **e**, Scatter plots for the CSF Aβ<sub>1-42</sub> level and PET SUVR/BeCKeT values. **f**, ROC analyses

coefficient for the overall data set (NCGG + AIBL, all tracers) was  $r = 0.678$  ( $P < 0.0001$ ). There were no significant correlations between the biomarker values and age or gender in the overall data set but a correlation between the composite biomarker and APOE4 ( $r = 0.464$ ,  $P < 0.0001$ ) was observed, and the partial correlation adjusted for SUVR/BeCKeT (standardized uptake value ratio, before the centiloid kernel transformation<sup>17</sup>) was  $r = 0.247$  ( $P < 0.0001$ ).

To further investigate the topographical associations between the plasma biomarkers and brain Aβ deposition, we conducted regression analyses using SPM8 software (see Methods). The results showed significant and robust correlations between the plasma biomarkers and areas of high Aβ deposition in the brain. The best association was

among Aβ-PET, CSF Aβ<sub>1-42</sub> and the plasma biomarkers. Data are from 46 individuals. ROC analyses of the plasma composite biomarker and CSF Aβ<sub>1-42</sub> to Aβ-PET (top). The performances of the composite biomarker and CSF Aβ<sub>1-42</sub> are AUC = 83.8% and 87.4%, sensitivity = 80.0% and 64.0%, specificity = 81.0% and 100%, and accuracy = 80.4% and 80.4%, respectively. ROC analyses of the plasma biomarkers to CSF Aβ<sub>1-42</sub>, using a standard determinant for Aβ-positivity with a cut-off value of 544 ng l<sup>-1</sup> (bottom). The composite biomarker showed AUC = 87.6%, sensitivity = 100%, specificity = 69.0%, and accuracy = 80.4%. For the scatter plots in **a**, **d**, and **e**, the coloured circles represent clinical categories: AD (red), MCI (orange) and cognitively normal (blue). Pearson's correlation coefficients ( $r$ ) and their significance (two-sided  $P$ ) are presented in the plots. The vertical dashed lines represent the cut-off values of each Aβ-PET imaging tracer (**a**) and CSF Aβ<sub>1-42</sub> (544 ng l<sup>-1</sup>) (**d**, **e**). Horizontal dashed lines represent the common cut-off values of the plasma biomarkers estimated in Extended Data Fig. 7a, d and of SUVR/BeCKeT (1.4) (**e**).

observed when using the composite biomarker in both unadjusted and adjusted (age and APOE4) analyses (Fig. 3b). The topographical association patterns were similar both in APOE4-positive and -negative sub-group analyses (Fig. 3c). These results demonstrate the strong association between the plasma biomarkers and Aβ deposition in the brain.

We also analysed the relationships between plasma biomarkers and CSF Aβ<sub>1-42</sub>. All of the plasma biomarkers, including the plasma Aβ<sub>1-42</sub> peptide alone, showed significant correlations with CSF Aβ<sub>1-42</sub> concentrations in an AIBL sub-group ( $n = 46$ ) (Fig. 3d). The composite biomarker demonstrated the highest correlation ( $r = -0.660$ ,  $P < 0.0001$ ) with CSF Aβ<sub>1-42</sub>, which is as high as the correlation between the CSF Aβ<sub>1-42</sub> and Aβ-PET SUVR/BeCKeT ( $r = -0.698$ ,  $P < 0.0001$ ) (Fig. 3e).

In order to further elucidate the relevance of the three different types of A $\beta$ -related biomarkers, we conducted ROC analyses among A $\beta$ -PET, CSF A $\beta_{1-42}$ , and plasma biomarkers. If A $\beta$ -PET is used as the standard classifier for A $\beta^+$ /A $\beta^-$  status, the plasma composite biomarker and CSF A $\beta_{1-42}$  showed identical accuracy (80.4%) with AUCs 83.8% and 87.4%, respectively (Fig. 3f, upper). Also, if we use the CSF A $\beta_{1-42}$  as the standard classifier, the plasma composite biomarker showed 87.6% AUC and 80.4% accuracy (Fig. 3f, lower). The performance of the plasma A $\beta$  composite biomarker was comparable to that of CSF A $\beta$  biomarkers<sup>18–22</sup>. These results demonstrate that the three different types of A $\beta$ -related biomarker (plasma and CSF A $\beta$ , and PET imaging), are highly correlated with each other, clearly indicating that plasma A $\beta$  biomarkers are strongly linked with the A $\beta$  status of the CNS, but less affected by the A $\beta$  known to be produced in peripheral tissues<sup>23</sup>.

The reasons for the high performance of the plasma A $\beta$  assays and the reliability of our IP-MS method are discussed in detail in the Supplementary Discussion and demonstrated in Extended Data Fig. 5. It should be reiterated that our biomarkers are not peptide levels, but are the ratios of plasma A $\beta_{1-42}$  to the reference peptides APP<sub>669–711</sub> and A $\beta_{1-40}$ . As these reference peptides have similar amino acid sequences and molecular sizes to A $\beta_{1-42}$ , the large inter-individual variances in plasma A $\beta$  levels, which are influenced by a wide variety of conditions<sup>5,23,24</sup> or anti-A $\beta$  autoantibodies<sup>1</sup>, should be reduced by the use of ratios. Several reports have proposed that the plasma ratio of A $\beta_{1-40}$  and A $\beta_{1-42}$  could be useful as a surrogate for brain A $\beta$  status, although its performance has not been sufficient to allow reliable prediction of the individual status of brain A $\beta$  burden<sup>5,6,25</sup>. A $\beta_{1-40}$  is known to aggregate less than A $\beta_{1-42}$ <sup>26</sup>, but neither the nature nor the molecular behaviour of APP<sub>669–711</sub> is known. Therefore, we performed two *in vitro* experiments, and found that APP<sub>669–711</sub> is a real neuronal product (Supplementary Information and Extended Data Fig. 6a) and that APP<sub>669–711</sub> showed much less self-assembly tendency than A $\beta_{1-42}$  (Supplementary Information and Extended Data Fig. 6b–d).

There were considerable site differences in plasma A $\beta$  and biomarker levels between the NCGG and AIBL data sets. We speculate that these were mainly due to pre-analytic factors, for example, differences in the procedures used for plasma processing (Supplementary Discussion). These between-site differences may complicate the establishment of a common cut-off value, which is essential for the widespread and multicentre use of biomarkers. It should be noted that this problem is still the biggest issue for CSF biomarkers, as highlighted recently<sup>27</sup>, and it is proving difficult to solve. To elucidate the influence of these between-site differences on biomarker performance, we explored optimal common cut-off values applicable for both sites by performing additional ROC analyses for the combined data sets (NCGG + AIBL), which allowed us to assess the classifying ability under the same biomarker levels across the sites. The results demonstrated that the biomarker performances were also high in the combined data sets (Extended Data Fig. 7a), and that the composite biomarker showed the highest classifying ability (Extended Data Fig. 7b). When we applied cut-off values determined by the Youden's index (see Methods) in the NCGG + AIBL overall data set, the performances were still very high in both the discovery and validation data sets (Extended Data Fig. 7c); for example, the composite biomarker showed 87.6% and 87.4% accuracy in the NCGG PIB and AIBL PIB data sets, respectively. The relevance of the cut-off values for biomarker performances is further visualized by the diagnostic performance plots in Extended Data Fig. 7d. These results support the stability of the biomarker performances with identical cut-off levels between sites.

Finally, we estimated possible clinical utility of the plasma biomarkers in several practical situations. First, we assessed the potential benefit of the plasma composite biomarker assuming two specific settings: screening for preclinical AD or prodromal AD to identify potential clinical trial candidates (see Supplementary Discussion and Extended Data Fig. 8a, b). Both scenarios suggest that the plasma biomarker screens could reduce unnecessary A $\beta$ -PET scans, substantially facilitating

recruitment for clinical trials. Furthermore, we assessed the potential utility of the plasma biomarker in daily clinical practice. When there is diagnostic uncertainty about a clinical diagnosis of AD, A $\beta$ -PET is considered to have a major clinical effect, providing diagnostic confidence or leading to changes in diagnosis<sup>28</sup>. In the NCGG data set, there were 9 out of 29 (31%) patients who had been diagnosed with AD but were PIB-PET negative, and the composite biomarker classified eight of them as A $\beta$  negative; therefore, the plasma biomarkers can also be expected to play an important clinical role. To confirm this possibility, we conducted an additional study with a new clinical data set consisting of 31 AD (22 A $\beta^+$  and 9 A $\beta^-$ , classified by PIB-PET) and 20 non-AD (8 A $\beta^+$  and 12 A $\beta^-$ ) cases (see Supplementary Discussion and Extended Data Fig. 8c, d). The plasma composite biomarker showed 96.7% sensitivity, 81.0% specificity, and 90.2% accuracy in the overall data ( $n = 51$ ) when predicting individual A $\beta$  status (A $\beta^+$  or A $\beta^-$ ) using the common cut-off value (0.376) (Extended Data Fig. 8e–g). The results suggest that the plasma biomarker could be helpful for the differential diagnosis of AD and aid in determining therapeutic strategies, by providing additional information on the brain A $\beta$  deposition status of individuals. As cost-benefit analysis of the use of A $\beta$ -PET for this purpose has proven controversial<sup>29</sup>, the impact of the plasma biomarker on daily clinical practice could be substantial.

The findings of the present study are considered to be robust, reproducible and reliable because biomarker performance was validated in a blinded manner using independent data sets (Japan and Australia) and involved an established large-scale multicentre cohort (AIBL). However, there are still several issues that need to be addressed before general clinical application can be considered. First, further validation studies (preferably in subjects drawn from primary care settings) coupled with longitudinal data will be needed. Second, standardized operating procedures for the analytical process as well as the pre- and post-analytical steps should be established<sup>30</sup>, preferably through an international consortium. Under the controlled and standardized operating procedures, optimal cut-off values as well as the optimal mathematical generation of the composite biomarker (see Supplementary Discussion and Extended Data Fig. 9a–c) should be established. Third, in clinical trials targeting A $\beta$  reduction, the usefulness of this plasma A $\beta$  biomarker as a monitoring tool remains to be evaluated. Fourth, biomarker performances for the differential diagnosis of other types of dementia need to be established. Finally, development of an automated assay system to stabilize the analytic factor and to enhance throughput of the IP-MS method is underway.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 March; accepted 11 December 2017.**

**Published online 31 January 2018.**

- Sevigny, J. *et al.* The antibody aducanumab reduces A $\beta$  plaques in Alzheimer's disease. *Nature* **537**, 50–56 (2016).
- Sperling, R., Mormino, E. & Johnson, K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron* **84**, 608–622 (2014).
- Henriksen, K. *et al.* The future of blood-based biomarkers for Alzheimer's disease. *Alzheimers Dement.* **10**, 115–131 (2014).
- O'Bryant, S. E. *et al.* Blood-based biomarkers in Alzheimer disease: Current state of the science and a novel collaborative paradigm for advancing from discovery to clinic. *Alzheimers Dement.* **13**, 45–58 (2017).
- Rembach, A. *et al.* Changes in plasma amyloid  $\beta$  in a longitudinal study of aging and Alzheimer's disease. *Alzheimers Dement.* **10**, 53–61 (2014).
- Swaminathan, S. *et al.* Association of plasma and cortical amyloid beta is modulated by APOE  $\epsilon 4$  status. *Alzheimers Dement.* **10**, e9–e18 (2014).
- Lövheim, H. *et al.* Plasma concentrations of free amyloid  $\beta$  cannot predict the development of Alzheimer's disease. *Alzheimers Dement.* **13**, 778–782 (2017).
- Kaneko, N. *et al.* Novel plasma biomarker surrogating cerebral amyloid deposition. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **90**, 353–364 (2014).
- Ovod, V. *et al.* Amyloid  $\beta$  concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis. *Alzheimers Dement.* **13**, 841–849 (2017).
- Gelfanova, V. *et al.* Quantitative analysis of amyloid- $\beta$  peptides in cerebrospinal fluid using immunoprecipitation and MALDI-ToF mass spectrometry. *Brief. Funct. Genomic. Proteomic.* **6**, 149–158 (2007).

11. Kaneko, N., Yamamoto, R., Sato, T. A. & Tanaka, K. Identification and quantification of amyloid  $\beta$ -related peptides in human plasma using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **90**, 104–117 (2014).
12. Ellis, K. A. *et al.* The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int. Psychogeriatr.* **21**, 672–687 (2009).
13. Vandenberghe, R. *et al.*  $^{18}\text{F}$ -flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann. Neurol.* **68**, 319–329 (2010).
14. Wong, D. F. *et al.* *In vivo* imaging of amyloid deposition in Alzheimer disease using the radioligand  $^{18}\text{F}$ -AV-45 (florbetapir F 18). *J. Nucl. Med.* **51**, 913–920 (2010).
15. Landau, S. M. *et al.* Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. *Eur. J. Nucl. Med. Mol. Imaging* **41**, 1398–1407 (2014).
16. Mormino, E. C. *et al.* Amyloid and APOE  $\epsilon 4$  interact to influence short-term decline in preclinical Alzheimer disease. *Neurology* **82**, 1760–1767 (2014).
17. Villemagne, V. L. *et al.* En attendant centiloid. *Adv. Res.* **2**, 723–729 (2014).
18. Fagan, A. M. *et al.* Comparison of analytical platforms for cerebrospinal fluid measures of  $\beta$ -amyloid 1-42, total tau, and p-tau181 for identifying Alzheimer disease amyloid plaque pathology. *Arch. Neurol.* **68**, 1137–1144 (2011).
19. Irwin, D. J. *et al.* Comparison of cerebrospinal fluid levels of tau and A $\beta$  1-42 in Alzheimer disease and frontotemporal degeneration using 2 analytical platforms. *Arch. Neurol.* **69**, 1018–1025 (2012).
20. Jagust, W. J. *et al.* Relationships between biomarkers in aging and dementia. *Neurology* **73**, 1193–1199 (2009).
21. Li, Q. X. *et al.* Alzheimer's disease normative cerebrospinal fluid biomarkers validated in PET amyloid- $\beta$  characterized subjects from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. *J. Alzheimers Dis.* **48**, 175–187 (2015).
22. Shaw, L. M. *et al.* Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann. Neurol.* **65**, 403–413 (2009).
23. Wang, J., Gu, B. J., Masters, C. L. & Wang, Y. J. A systemic view of Alzheimer disease - insights from amyloid- $\beta$  metabolism beyond the brain. *Nat. Rev. Neurol.* **13**, 612–623 (2017).
24. Janelidze, S. *et al.* Plasma  $\beta$ -amyloid in Alzheimer's disease and vascular disease. *Sci. Rep.* **6**, 26801 (2016).
25. Rembach, A. *et al.* Plasma amyloid- $\beta$  levels are significantly associated with a transition toward Alzheimer's disease as measured by cognitive decline and change in neocortical amyloid burden. *J. Alzheimers Dis.* **40**, 95–104 (2014).
26. Jarrett, J. T., Berger, E. P. & Lansbury, P. T., Jr. The C-terminus of the  $\beta$  protein is critical in amyloidogenesis. *Ann. NY Acad. Sci.* **695**, 144–148 (1993).
27. Rogers, M. B. Are CSF Assays Finally Ready for Prime Time? *Alzforum* <https://www.alzforum.org/news/conference-coverage/are-csf-assays-finally-ready-prime-time> (2017).
28. Boccardi, M. *et al.* Assessment of the incremental diagnostic value of florbetapir F 18 imaging in patients with cognitive impairment: the incremental diagnostic value of amyloid PET with [ $^{18}\text{F}$ ]-florbetapir (INDIA-FBP) Study. *JAMA Neurol.* **73**, 1417–1424 (2016).
29. Caselli, R. J. & Woodruff, B. K. Clinical impact of amyloid positron emission tomography—is it worth the cost? *JAMA Neurol.* **73**, 1396–1398 (2016).
30. O'Bryant, S. E. *et al.* Guidelines for the standardization of preanalytic variables for blood-based biomarker studies in Alzheimer's disease research. *Alzheimers Dement.* **11**, 549–560 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The NCGG study group thank all participants of the study, clinicians who referred patients, and all the staff who supported the MULNIAD project. We thank N. Sugimoto for conducting statistical analyses, S. Niida and the NCGG Biobank members for the management of plasma samples and M. Kawakage for data monitoring. This study was supported by The Research Funding for Longevity Sciences (25-24 and 26-30) from the National Center for Geriatrics and Gerontology, and partially supported by Research and Development Grants for Dementia from the Japan Agency for Medical Research and Development, AMED. This study is registered under UMIN ID: 000016144. The AIBL study would like to thank all participants of the study and the clinicians who referred participants. The AIBL study (<https://aibl.csiro.au/>) is a consortium between Austin Health, CSIRO, Edith Cowan University and the Florey Institute, The University of Melbourne. Partial financial support was provided by the Alzheimer's Association (US), the Alzheimer's Drug Discovery Foundation, an anonymous foundation, the Cooperative Research Centre for Mental Health, CSIRO Science and Industry Endowment Fund, the Dementia Collaborative Research Centres, the Victorian Government Operational Infrastructure Support program, the McCusker Alzheimer's Research Foundation, the National Health and Medical Research Council, and the Yulgilbar Foundation. Funding sources had no role in study design, data collection, data analyses or data interpretation.

**Author Contributions** A.N., N.K., T.K., K.It., K.T., and K.Y. designed the study. A.N., N.K., V.L.V., C.L.M., and K.Y. wrote the manuscript and A.N., N.K., V.L.V., V.D., T.T., and K.M. made the figures. A.N., V.L.V., T.K., V.D., C.F., Q.-X.L., R.M., C.R., Ke.Is., Ka.Is., Y.A., and C.L.M. collected the data. A.N., N.K., V.L.V., T.K., J.D., V.D., C.F., Q.-X.K., R.M., C.R., Y.A., T.T., K.M., S.I., K.It., K.T., and C.L.M. analysed the data. All authors interpreted the data and critically revised the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to K.Y. ([katuhiko@nccg.go.jp](mailto:katuhiko@nccg.go.jp)).

**Reviewer Information** *Nature* thanks H. Federoff, R. Thomas and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Subjects.** The participants were aged 60 to 90 years. The discovery NCGG data set consisted of 62 cognitively normal individuals, 30 with MCI, and 29 with AD (121 in total) selected from in-house clinical studies at the NCGG. The AIBL data set for external validation consisted of 156 cognitively normal individuals, 68 with MCI, and 30 with AD participants (254 in total).

All participants from the NCGG were native Japanese, recruited from community dwellings and outpatients of the National Hospital for Geriatric Medicine at NCGG. The clinical classification of NCGG subjects was determined by following the inclusion criteria of the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) study (<http://adni.loni.usc.edu/>). The definition of the cognitively normal group in the NCGG data set is generally equivalent to the cognitively normal group in the ADNI2 study. All of the AD and MCI subjects also fulfilled the diagnostic criteria developed by the National Institute on Aging and the Alzheimer's Association (NIA-AA)<sup>31,32</sup>. The samples were selected on the basis of age, clinical category (cognitively normal, MCI or AD), and data availability for both plasma measurements and A $\beta$ -PET imaging data. Subjects under treatment for any substantial medical, neurological, or psychiatric disease, or with any history of a major psychiatric disorder, alcohol dependence, or substance dependence, were excluded. Individuals with any clinically significant focal brain lesion by MRI were also excluded. There were no individuals at the extremes of socio-economic status.

AIBL is a two-site (Melbourne and Perth), longitudinal cohort study, integrating neuroimaging, biomarker, neuropsychometric, and lifestyle data. The AIBL study population was selected from English-speaking volunteers who responded to media advertisements, or clinical cases that were referred to the study by a network of doctors. The AIBL study has strict selection criteria to eliminate, as much as possible, comorbidities such as vascular disease and diabetes, but no requirement on socio-economic status. Approximately 48% of the AIBL cohort reported more than 13 years of education. Clinical classification of the AIBL study was determined as previously described<sup>12</sup>. The AIBL samples were selected with the same conditions as those selected from the NCGG, so that age, sex, and clinical category matched.

In both the NCGG and AIBL data sets, all selected subjects had stored plasma samples and corresponding A $\beta$ -PET imaging data that were acquired within one year of plasma sampling. The mean and s.d. of the time discrepancies between plasma sampling and PET imaging were  $41.1 \pm 97.5$  and  $115.7 \pm 93.9$  days for NCGG and AIBL, respectively.

Both studies were approved by the appropriate institutional ethics committee (NCGG Ethics Committee, Japan, and Human Research Ethics Committee, Research Governance Unit, St Vincent's Healthcare, Australia, respectively), and were performed following all relevant ethical regulations. Written informed consent was obtained from all participants (or their legal guardians) before participation.

From the 254 plasma samples in the total AIBL data set, two outliers were excluded from the analyses. One subject's abnormally high A $\beta$  signals from IP-MS masked the peak of the internal standard which prevented reliable measurements, and the other subject showed A $\beta$  concentrations 9.2–20.5 times higher than the s.d.

**Imaging data.** A $\beta$ -PET imaging for the discovery set in NCGG was performed with <sup>11</sup>C-PIB (PIB), while A $\beta$  imaging for the AIBL validation set was performed with three different radiotracers: PIB, FLUTE, or FBP. The PET methodology for each tracer has been previously described<sup>33</sup>. In brief, PET images were spatially normalized with CapAIBL using an adaptive atlas<sup>34</sup>, and sampled using a preset template of narrow cortical regions of interest (ROI). For semi-quantitative analysis, a volume of interest template was applied to the summed and spatially normalized PET images in order to obtain a standardized uptake value (SUV). The images were then scaled to the SUV of each tracer's recommended reference region to generate a tissue ratio termed the SUV ratio (SUVR). A global measure of A $\beta$  burden was computed using the mean SUVR in the frontal, superior parietal, lateral temporal, lateral occipital, and anterior and posterior cingulate regions. For PIB, the SUVs were normalized to the cerebellar cortex, the whole cerebellum was used as the reference region for FBP<sup>35</sup> while for FLUTE the pons was used as the reference region<sup>36</sup> as advocated by the pharmaceutical companies that supplied each tracer. The SUVR was dichotomized as having a high (A $\beta^+$ ) or low (A $\beta^-$ ) A $\beta$  burden, using a cut-off value that was determined for each tracer. Participants who underwent PIB were considered to have high A $\beta$  when SUVR  $\geq 1.40$ , for FLUTE when SUVR  $\geq 0.55$  and for FBP when SUVR  $\geq 1.05$ . For the analysis across different PET tracers, BeCKET values, which are a linear transformed standardization of FLUTE and FBP SUVR onto 'PIB-like' SUVR<sup>17</sup>, were used.

Individual MRI with high-resolution 3D T1-weighted, T2-weighted, and fluid-attenuated inversion recovery images were acquired for all NCGG and AIBL participants. These MRIs were used to exclude subjects who had substantial brain lesions.

Voxelwise-correlational analyses were performed after spatially normalizing and scaling all PET images with CapAIBL<sup>34</sup>. In brief, a combined plasma-PET statistical analysis was performed using SPM8 software, in which the associations between plasma biomarkers and A $\beta$ -PET were estimated using a linear regression model within each cohort (discovery, validation, and all PIB combined) and each tracer (PIB, FLUTE, FBP, and all combined), using one plasma biomarker at a time as the dependent variable and A $\beta$ -PET as the independent variable. The models were further examined after adjusting for age and APOE status. The statistical threshold for the voxelwise computations in SPM8 was set at  $P < 0.05$ , using FWE to correct for multiple comparisons at a peak level.

**Blood processing and plasma storage.** In the NCGG study, blood samples were collected between 11:00 and 15:00. Plasma was isolated from whole blood collected in 7-ml EDTA-2Na tubes (Venject II, TERUMO). Within 5 min of blood collection, whole blood was centrifuged at 2273g for 5 min at room temperature. Otherwise, the blood was temporally stored on ice for up to 30 min, and then centrifuged. The plasma was immediately transferred to storage tubes (48 Jacket Tubes 2.0ml External-Type, FCR&Bio) as 250- or 500- $\mu$ l aliquots, and frozen immediately in a  $-80^\circ\text{C}$  freezer. The plasma samples were stocked in the NCGG biobank, where each was assigned its own ID independent of the study ID. The research group did not intervene in sample collection and shipping to the Shimadzu Corporation for IP-MS assays was performed in a blinded manner.

In AIBL, blood samples were collected between 9:00 and 10:00. Plasma was isolated from whole blood collected in Sarstedt s-monovette 7.5-ml EDTA tubes (Sarstedt) with pre-added Prostaglandin E1 (PGE1, Sapphire Bioscience) to produce a final PGE1 concentration of  $33 \text{ ng ml}^{-1}$  of whole blood. Processing started after bloods had equilibrated with room temperature and within 1 h of collection. Whole blood was centrifuged at 200g for 10 min at room temperature (no deceleration) to generate a platelet-rich plasma (PRP) layer. The PRP was transferred using 3-ml transfer pipettes (Livingstone) to a new 15-ml polypropylene centrifuge tube (Greiner Bio-One CELLSTAR). Both the collection tube and 15-ml tubes were centrifuged at 800g for 15 min at room temperature, maximum deceleration. Plasma was combined into a new 15-ml polypropylene tube and spun at 3,200g for 30 min to remove debris. Plasma, as 250- $\mu$ l aliquots, was stored in 1-ml capacity NUNC 2D barcoded Bank-IT polypropylene cryovials (NUNC) and frozen immediately on dry ice before long term storage in vapour-phase liquid nitrogen. For the AIBL samples, following confirmation that the proposed cohort had age, gender and APOE4 matching between the clinical groups, a new study ID was attached to the blood tubes before shipment to Shimadzu. Researchers at Shimadzu were blinded to any associated clinical or PET data until the data collected were complete and locked.

**Plasma A $\beta$  measurements.** Plasma A $\beta$  levels were measured using IP-MS, which is an analytical technique that quantifies A $\beta$ -related peptides of different mass in MALDI-TOF mass spectrometry after they have been isolated and enriched from abundant plasma proteins by immunoprecipitation using the specific affinity of an antibody. This assay was modified from previously reported procedures<sup>8</sup> with two major modifications being made. First, general antibody beads, prepared by coupling intact IgG monoclonal antibody 6E10 (BioLegend) directly to Dynabeads M-270 Epoxy (Thermo Fisher Scientific) according to the manufacturer's protocol, were used for immunoprecipitation. This method for preparing the antibody beads is more simple and practical than the previously reported method because it does not require generation and purification of two antigen-binding fragments (Fab') from IgGs (clone 6E10 and 4G8) or the coupling of them on beads through PEG. Second, the immunoprecipitation procedure was carried out using two rounds of repeated processing, which both reduces the non-specific binding of abundant proteins that interferes with the signals of the A $\beta$ -related peptides, and increases specificity during the detection of A $\beta$ -related peptides in MALDI-TOF mass spectrometry.

In detail, 250  $\mu$ l of plasma was mixed with an equal volume of Tris buffer containing 10 pM stable-isotope-labelled (SIL) A $\beta_{1-38}$  peptide (AnaSpec, San Jose, CA), 0.2% w/v *n*-dodecyl- $\beta$ -D-maltoside (DDM) and 0.2% w/v *n*-nonyl- $\beta$ -D-thiomaltoside (NTM). The SIL-A $\beta_{1-38}$  peptide was used as internal standard for normalization of signals for all A $\beta$ -related peptides in the mass spectrum, which was different from other mass spectrometry-based studies<sup>9,37,38</sup> that used synthetic peptides corresponding to each A $\beta$ -related peptide as the internal standards (for example, SIL-A $\beta_{1-42}$  for A $\beta_{1-42}$ , and SIL-A $\beta_{1-40}$  for A $\beta_{1-40}$ ). This was because A $\beta_{1-38}$  is relatively easy to deal with as it has a lower self-aggregating tendency and lower adsorption in storage tubes when compared to A $\beta_{1-42}$ <sup>39-41</sup>. More importantly, using a common internal standard has an advantage when computing peptide ratios, because it can cancel out any implicit errors related to the amounts of added SIL-A $\beta_{1-38}$  caused by activities such as production, preparation and/or handling. Furthermore, using only one standard peptide is simpler than handling three standard peptides, which has cost-benefit implications (see Supplementary

Discussion and Extended Data Fig. 5d–f). The nonionic detergents DDM and NTM were used for reducing nonspecific binding and obtaining high signals of A $\beta$ -related peptides in MALDI–TOF mass spectrometry. The plasma A $\beta$ -related peptides and internal standard were immunoprecipitated by incubating the antibody beads with the plasma sample for 1 h. The bound peptides were washed and eluted with glycine buffer (pH 2.8) containing 0.1% w/v DDM. After the pH was adjusted to 7.4 with Tris buffer, the immunoprecipitation was repeated once and the bound peptides were eluted with 70% acetonitrile containing 5 mM HCl. The eluted peptides were applied on four wells of a 900- $\mu$ m Focus MALDI plate<sup>TM</sup> (Hudson Surface Technology) which was prespotted with  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) and methanediphosphonic acid (MDPNA). Mass spectra were acquired using a MALDI-linear TOF mass spectrometer (AXIMA Performance, Shimadzu/KRATOS) equipped with a 337-nm nitrogen laser in the positive ion mode. The  $m/z$  value and signal variability in the mass spectrometer were calibrated externally with a mixture of standard peptides to improve the precision of the A $\beta$ -related peptide signal peak. The peak intensities were extracted using Mass++ software v2 (ref. 42) (Shimadzu). The peptide mass tolerance for quantification was set within 2.5 Da of the theoretical mass. The limit of detection was established at a signal-to-noise ratio of 3:1. One assay produced four mass spectra and the levels of plasma A $\beta$ -related peptides were obtained by averaging the four spectra normalized with SIL-A $\beta_{1-38}$ . The normalized intensity was used as plasma A $\beta$ -related peptide levels. The quantitiveness and reliability of the IP–MS assay were carefully validated by several steps as detailed in the Supplementary Discussion and Extended Data Fig. 5a–c. Using the IP–MS method, we tested linear relationships between the normalized signal intensity and the concentration of A $\beta$ -related peptides in PBS containing 3 mg ml<sup>−1</sup> bovine serum albumin, and in human plasma, and ensured the reliability of quantification. For example, we analysed the dose dependency of the normalized intensity for each of three synthetic peptides (A $\beta_{1-42}$ , A $\beta_{1-40}$ , and APP<sub>669-711</sub>) that had been spiked into the human plasma. The results showed very good linearity for each peptide with coefficients of determination ( $R^2$ ) between 0.999 and 1.000 (Extended Data Fig. 5c). These  $R^2$  values are as high as, or even higher than, those reported in a mass spectrometry-based study that used SIL-A $\beta_{1-42}$  and SIL-A $\beta_{1-40}$  as internal standards for each corresponding peptide measured in CSF<sup>37</sup>. Each peptide is ionized differently during mass spectrometry; our standard curves show different slopes as previously reported<sup>37</sup>, but this does not affect the robustness and reproducibility of quantification. We also verified the reproducibility of the assay using human EDTA plasma (Tennessee Blood Services). The intra- and inter-day assay coefficients of variance obtained for A $\beta_{1-40}$  were 4.2–4.7% ( $n = 5$ ) and 3.2–6.8% ( $n = 3$ ), respectively; for A $\beta_{1-42}$  the coefficients were 6.8–7.8% and 1.6–7.7%, respectively, and for APP<sub>669-711</sub> the coefficients were 2.9–8.2% and 4.7–10.7%, respectively. These values are smaller than those obtained for within-laboratory CSF biomarker assays<sup>43</sup> (5% to 19%), supporting the reliability of our measurements. The IP–MS method can also measure other forms of plasma A $\beta$  such as A $\beta_{1-38}$  and A $\beta_{1-39}$ , but we did not focus on them in this study.

**CSF biomarker measurements.** In the AIBL data set, 46 subjects underwent CSF testing within two months of blood sampling and A $\beta$ -PET imaging. The procedures for CSF sampling and biomarker measurements were performed as previously described<sup>21</sup>. In this study, we focused on analysing CSF A $\beta_{1-42}$  values, which were measured by ELISA<sup>21</sup>. The cut off value of CSF A $\beta_{1-42}$  was 544 ng l<sup>−1</sup> (ref. 21, below which A $\beta_{1-42}$  was considered abnormal).

**Sample size considerations.** The power calculations for sample sizes in the study were estimated as follows: assuming that the biomarker candidates could be used to classify individuals as A $\beta^+$  or A $\beta^-$  with a sensitivity of  $\geq 80\%$  and a theoretical sensitivity of 50%, the sample size required to achieve a statistical power of 80% at a 5% significance level would be 20 and 20 for both A $\beta^+$  and A $\beta^-$  groups. Also, assuming that the plasma biomarkers could show more than a 0.5 correlation coefficient ( $r$ ) to A $\beta$ -PET SUVR values or to CSF biomarker values, a total sample size of 21 would be required to achieve a statistical power of 80% at a 5% significance level. Both the NCGG and AIBL data sets, including the subpopulation with CSF data, satisfied these sample size requirements.

**Data analyses.** Data analyses were performed in a blind and independent manner. The plasma-A $\beta$  measurements were performed at Koichi Tanaka Mass Spectrometry Research Laboratory (Shimadzu) without any clinical or imaging information. All of the PET imaging data were analysed by the AIBL imaging group. The A $\beta$ -PET dichotomization (A $\beta^+$ /A $\beta^-$ ) and generation of SUVR were performed without any clinical or biomarker information. The NCGG group conducted statistical analyses, and all results were confirmed by two independent biostatisticians. The test biomarker values were generated by computing the ratio of normalized intensity of A $\beta_{1-42}$  with APP<sub>669-711</sub>, and A $\beta_{1-40}$ . We used A $\beta_{1-42}$  as the denominator, because these ratios (APP<sub>669-711</sub>/A $\beta_{1-42}$  and A $\beta_{1-40}$ /A $\beta_{1-42}$ ) showed a normal distribution without any transformation in both the NCGG and AIBL data sets (Shapiro–

Wilk test<sup>44</sup>), whereas using A $\beta_{1-42}$  value as the numerator did not (Extended Data Fig. 2b). The composite biomarker was generated by combining the normalized scores of APP<sub>669-711</sub>/A $\beta_{1-42}$  and A $\beta_{1-40}$ /A $\beta_{1-42}$  as follows: first, the discovery NCGG data set was used for a standard database, and values of APP<sub>669-711</sub>/A $\beta_{1-42}$  and A $\beta_{1-40}$ /A $\beta_{1-42}$  in all data sets (both NCGG and AIBL) were normalized to  $z$ -scores using the mean (0.774 and 24.72, respectively) and s.d. (0.191 and 4.31, respectively) of the NCGG data; then,  $z$ -scores of APP<sub>669-711</sub>/A $\beta_{1-42}$  and A $\beta_{1-40}$ /A $\beta_{1-42}$  were averaged for each subject and used as a value for the composite biomarker so that each biomarker contributed equally to the composite. Before the main analyses, the weight of the  $z$ -score composition was pre-determined as 1:1 by exploratory analyses at NCGG that were confirmed by a pilot study.

Statistical analyses were performed using R v.3.3.2, SPSS v.21 (IBM), and JMP software v.8 (SAS Institute). For categorical data, such as gender, clinical category and APOE4 carrier distributions, group differences were analysed using the  $\chi^2$  test. For numerical data, group differences were analysed by Student's  $t$ -test or Welch's  $t$ -test, and the effect size was assessed using Cohen's  $d$ .

The biomarker performance when predicting A $\beta^+$ /A $\beta^-$  status was assessed using ROC analyses. The AUC, and the representative best values for the sensitivity, specificity and accuracy at an optimal cut-off point, were used for the performance measures. The cut-off points were determined by Youden's index<sup>45</sup>, which optimizes biomarker performance when equal weight is given to sensitivity and specificity. In addition, positive predictive value (PPV) and negative predictive value (NPV) were estimated by assuming the prevalence of A $\beta^+$  individuals in specific settings. These values were computed as follows:

Where TP = true positive, TN = true negative, FP = false positive, and FN = false negative; sensitivity = TP/(FN+TP), specificity = TN/(TN+FP), accuracy = (TP+TN)/(TP+TN+FP+FN), PPV = 1/(1 + ((1 − prevalence)/prevalence)((1 − specificity)/sensitivity)), and NPV = 1/(1 + (prevalence/(1 − prevalence))((1 − sensitivity)/specificity)).

We performed both unadjusted and adjusted ROC analyses. In unadjusted ROC analyses, original biomarker values from the discovery (NCGG) and validation (AIBL) data were used. In adjusted ROC analyses, a predictive formula including confounders (for example, age, gender, APOE4 and clinical category) was built using a generalized linear model (GLM) (binomial logistic regression analysis) on the discovery NCGG data as follows:

$$\pi = 1/(1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)})$$

Then the same formula and the same coefficients were applied to both the NCGG discovery and AIBL validation data to calculate the fitted predicted probabilities. These predictive values were used for the adjusted ROC analyses of the NCGG and AIBL data.

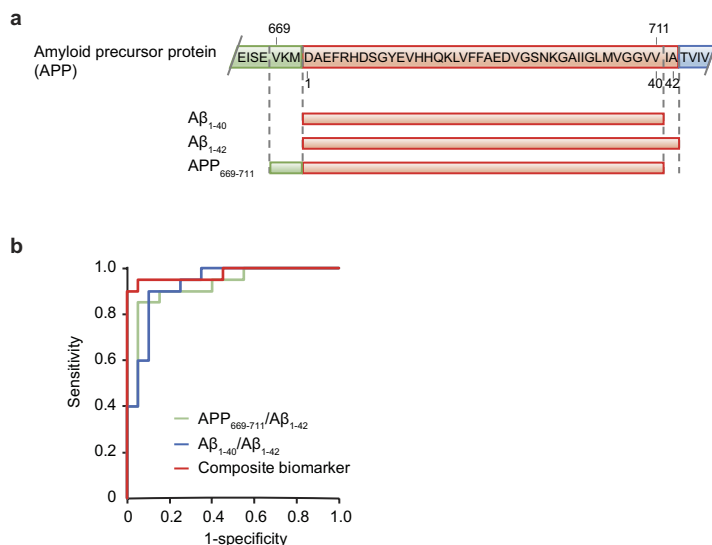
To compare the biomarker performances among A $\beta_{1-42}$ , APP<sub>669-711</sub>/A $\beta_{1-42}$ , A $\beta_{1-40}$ /A $\beta_{1-42}$ , and the composite biomarker within the same data set, the differences between pairs of AUCs were statistically analysed using the DeLong test<sup>46</sup>. The  $P$  values were Bonferroni-corrected by multiplying the  $P$  values by the number of comparisons (6) to control for the multiple comparisons problem. Improvement in the predictive ability of an alternative model was also assessed using the categorical NRI and IDI in the logistic regression model<sup>47</sup>. For the categorical NRI, the reclassification ability was measured in four categories using the first, second and third quartiles of the original model's fitted values as cut points. Statistical differences in AUCs between two different data sets were analysed using Delong's test for two uncorrelated ROC curves<sup>46</sup>.

Pearson product–moment correlational analysis was conducted to evaluate the strength of the association between each plasma biomarker and cortical-A $\beta$  deposition assessed by either A $\beta$ -PET imaging or CSF A $\beta$  values. All the tests were two-tailed, and the significance level of difference was set at  $P < 0.05$ .

**Data availability.** Source Data for graphs plotted in Figs 1–3 and Extended Data Figs 1–9 are available in the online version of this paper. All other data are available from the corresponding author upon reasonable request.

- Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 270–279 (2011).
- McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 263–269 (2011).
- Rowe, C. C. *et al.* Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiol. Aging* **31**, 1275–1283 (2010).
- Bourgeat, P. *et al.* Comparison of MR-less PiB SUVR quantification methods. *Neurobiol. Aging* **36**, S159–S166 (2015).

35. Clark, C. M. *et al.* Use of florbetapir-PET for imaging  $\beta$ -amyloid pathology. *J. Am. Med. Assoc.* **305**, 275–283 (2011).
36. Lundqvist, R. *et al.* Implementation and validation of an adaptive template registration method for  $^{18}\text{F}$ -flutemetamol imaging data. *J. Nucl. Med.* **54**, 1472–1478 (2013).
37. Pannee, J. *et al.* A selected reaction monitoring (SRM)-based method for absolute quantification of A $\beta$ 38, A $\beta$ 40, and A $\beta$ 42 in cerebrospinal fluid of Alzheimer's disease patients and healthy controls. *J. Alzheimers Dis.* **33**, 1021–1032 (2013).
38. Patterson, B. W. *et al.* Age and amyloid effects on human central nervous system amyloid-beta kinetics. *Ann. Neurol.* **78**, 439–453 (2015).
39. Manzoni, C. *et al.* Overcoming synthetic A $\beta$  peptide aging: a new approach to an age-old problem. *Amyloid* **16**, 71–80 (2009).
40. Schlenzig, D. *et al.* N-Terminal pyroglutamate formation of A $\beta$ 38 and A $\beta$ 40 enforces oligomer formation and potency to disrupt hippocampal long-term potentiation. *J. Neurochem.* **121**, 774–784 (2012).
41. Toombs, J., Paterson, R. W., Schott, J. M. & Zetterberg, H. Amyloid-beta 42 adsorption following serial tube transfer. *Alzheimers Res. Ther.* **6**, 5 (2014).
42. Tanaka, S. *et al.* Mass++: A visualization and analysis tool for mass spectrometry. *J. Proteome Res.* **13**, 3846–3853 (2014).
43. Mattsson, N. *et al.* CSF biomarker variability in the Alzheimer's Association quality control program. *Alzheimers Dement.* **9**, 251–261 (2013).
44. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
45. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
46. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
47. Pencina, M.J., D'Agostino, R.B. Sr., D'Agostino, R.B. Jr. & Vasan, R.S. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* **27**, 157–172 (2008).

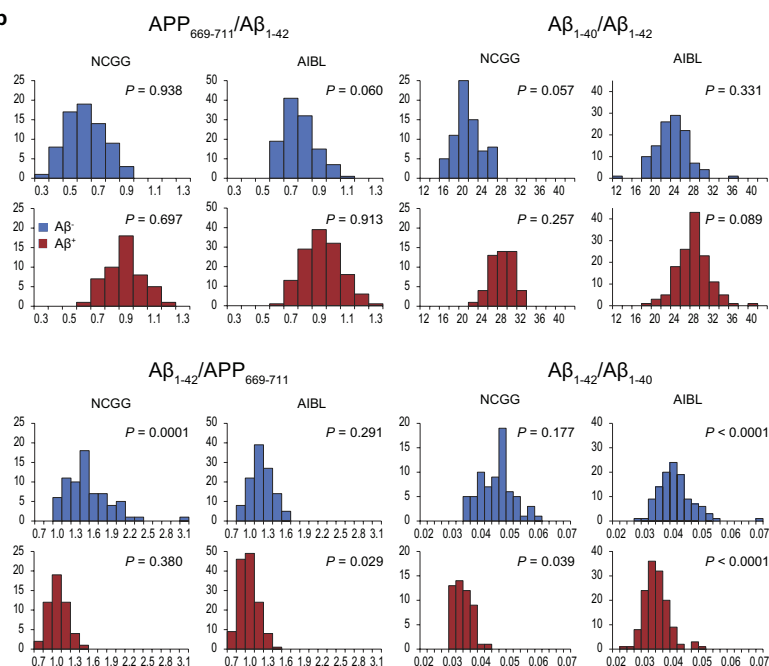


**Extended Data Figure 1 | The amino acid sequences of Aβ-related peptides and results of the pilot study. a,** Overview of the amino acid sequences of the Aβ-related peptides Aβ<sub>1-40</sub>, Aβ<sub>1-42</sub> and APP<sub>669-711</sub>. **b,** ROC analyses of the blinded pilot study for 20 Aβ<sup>+</sup> and 20 Aβ<sup>-</sup> subjects (see Supplementary Information).

The green, blue, and red curves indicate APP<sub>669-711</sub>/Aβ<sub>1-42</sub>, Aβ<sub>1-40</sub>/Aβ<sub>1-42</sub> and the composite biomarker, respectively. The AUCs and the representative best values of sensitivity, specificity and accuracy for these biomarkers as determined by Youden's index are as follows: APP<sub>669-711</sub>/Aβ<sub>1-42</sub>, AUC = 0.923, sensitivity = 0.850, specificity = 0.950, accuracy = 0.900; Aβ<sub>1-40</sub>/Aβ<sub>1-42</sub>, AUC = 0.930, sensitivity = 0.900, specificity = 0.900, accuracy = 0.900; composite biomarker, AUC = 0.975, sensitivity = 0.950, specificity = 0.950, accuracy = 0.950.

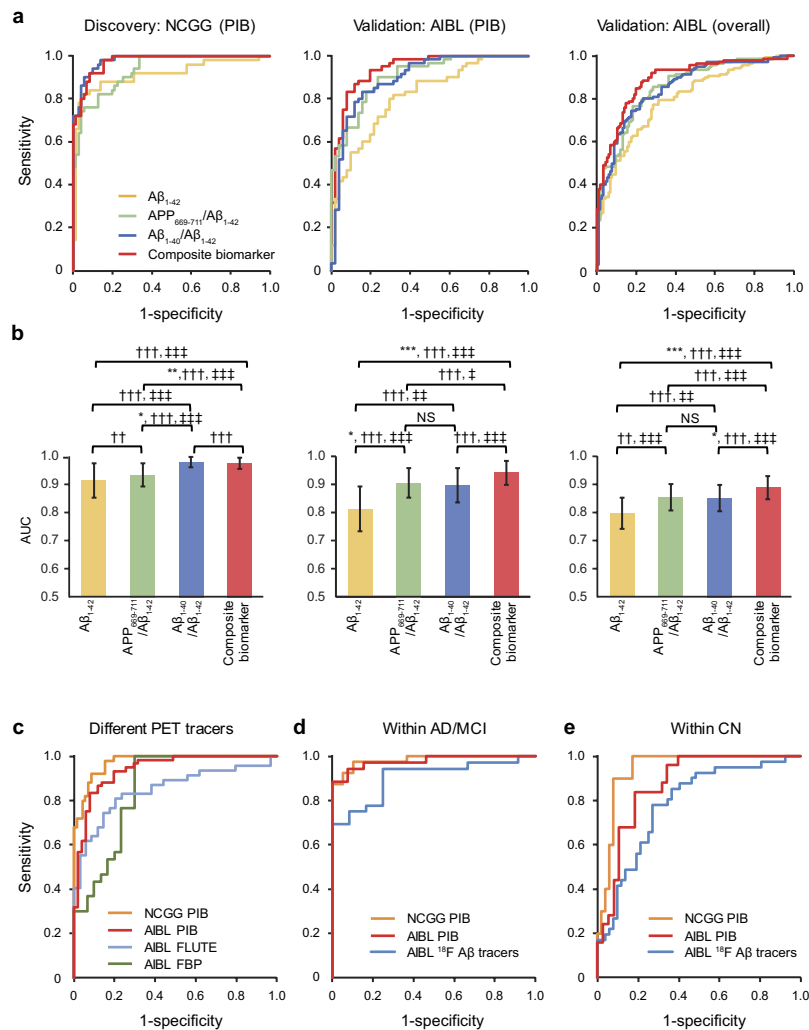
**a**The A $\beta$ -related peptide and biomarker values in each study sites (NCGG and AIBL)

| Peptides               | APP <sub>669-711</sub>                            |                | A $\beta$ <sub>1-40</sub>                            |                    | A $\beta$ <sub>1-42</sub> |                    |
|------------------------|---|----------------|--|--------------------|---------------------------|--------------------|
|                        | NCGG  | AIBL           | NCGG   | AIBL               | NCGG                      | AIBL               |
| overall                | 0.271   | 0.278          | 8.87   | 8.33 <sup>a</sup>  | 0.373                     | 0.316 <sup>c</sup> |
| (95% CI)               | (0.261, 0.282)                                    | (0.271, 0.285) | (8.50, 9.24)   | (8.10, 8.57)       | (0.352, 0.393)            | (0.307, 0.326)     |
| A $\beta$ <sup>+</sup> | 0.263   | 0.279          | 8.28   | 8.28               | 0.289                     | 0.291              |
| (95% CI)               | (0.246, 0.281)                                    | (0.270, 0.289) | (7.61, 8.94)   | (8.01, 8.55)       | (0.264, 0.314)            | (0.281, 0.301)     |
| A $\beta$ <sup>-</sup> | 0.277   | 0.277          | 9.29   | 8.39 <sup>b</sup>  | 0.431                     | 0.347 <sup>c</sup> |
| (95% CI)               | (0.265, 0.289)                                    | (0.265, 0.288) | (8.88, 9.69)   | (7.99, 8.79)       | (0.410, 0.453)            | (0.332, 0.362)     |
| <i>P</i> -value        | 0.174   | 0.708          | 0.011  | 0.651              | < 0.0001                  | < 0.0001           |
| Cohen's <i>d</i>       | 0.252   | 0.047          | 0.508  | 0.059              | 1.570                     | 0.791              |
| Biomarkers             | APP <sub>669-711</sub> /A $\beta$ <sub>1-42</sub> |                | A $\beta$ <sub>1-40</sub> /A $\beta$ <sub>1-42</sub> |                    | Composite biomarker       |                    |
|                        | NCGG  | AIBL           | NCGG   | AIBL               | NCGG                      | AIBL               |
| overall                | 0.774   | 0.896          | 24.72  | 26.70 <sup>c</sup> | 0                         | 0.548 <sup>c</sup> |
| (95% CI)               | (0.739, 0.808)                                    | (0.877, 0.914) | (23.95, 25.50)                                       | (26.21, 27.19)     | (-0.170, 0.170)           | (0.457, 0.640)     |
| A $\beta$ <sup>+</sup> | 0.934   | 0.971          | 28.84  | 28.69              | 0.896                     | 0.975              |
| (95% CI)               | (0.896, 0.972)                                    | (0.948, 0.993) | (28.19, 29.49)                                       | (28.13, 29.25)     | (0.755, 1.037)            | (0.871, 1.080)     |
| A $\beta$ <sup>-</sup> | 0.661   | 0.807          | 21.82  | 24.33 <sup>c</sup> | -0.631                    | 0.040 <sup>c</sup> |
| (95% CI)               | (0.629, 0.694)                                    | (0.786, 0.827) | (21.17, 22.48)                                       | (23.72, 24.93)     | (-0.776, -0.486)          | (-0.055, 0.135)    |
| <i>P</i> -value        | < 0.0001  | < 0.0001       | < 0.0001   | < 0.0001           | < 0.0001                  | < 0.0001           |
| Cohen's <i>d</i>       | 2.007   | 1.335          | 2.730  | 1.319              | 2.686                     | 1.630              |

**b**

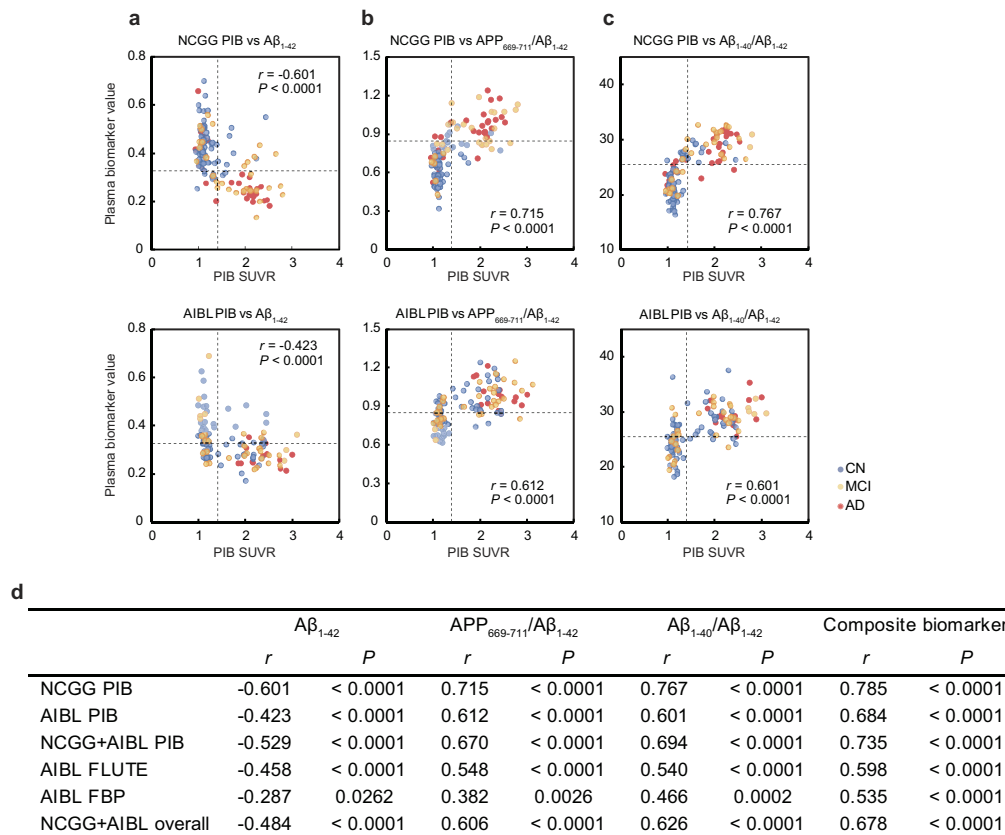
**Extended Data Figure 2 | Peptide and biomarker values, and their distributions. a,** The A $\beta$ -related peptide and biomarker values in each study site (NCGG and AIBL). Normalized intensity of each peptide (top), and values for each biomarker (bottom) in the NCGG ( $n = 121$ ) and AIBL overall ( $n = 252$ ) data sets. Composite biomarker values are the average of the normalized values of APP<sub>669-711</sub>/A $\beta$ <sub>1-42</sub> and A $\beta$ <sub>1-40</sub>/A $\beta$ <sub>1-42</sub>. Peptide values are arbitrary units. Means and 95% confidence intervals (CI) in parentheses; *P* values show statistical differences between the A $\beta$ <sup>+</sup> and A $\beta$ <sup>-</sup> groups (two-sided Student's *t*-test or Welch's *t*-test). Superscripts

indicate statistically significant site differences (<sup>a</sup> $P = 0.012$ , <sup>b</sup> $P = 0.002$ , <sup>c</sup> $P = < 0.0001$ , two-sided). These site differences did not change when using analysis of covariance (ANCOVA) adjusted for semi-quantitative measures of A $\beta$ -PET, using SUVR (PIB) and BeCKeT (FLUTE and FBP) values. **b,** Histograms of the biomarker value distributions for APP<sub>669-711</sub>/A $\beta$ <sub>1-42</sub>, A $\beta$ <sub>1-40</sub>/A $\beta$ <sub>1-42</sub> (top), and their inversions (A $\beta$ <sub>1-42</sub>/APP<sub>669-711</sub> and A $\beta$ <sub>1-42</sub>/A $\beta$ <sub>1-40</sub>) (bottom). Blue and red bars represent the distributions of A $\beta$ <sup>+</sup> and A $\beta$ <sup>-</sup> populations, respectively. *P* values represent the results of the Shapiro–Wilk test (see Methods).



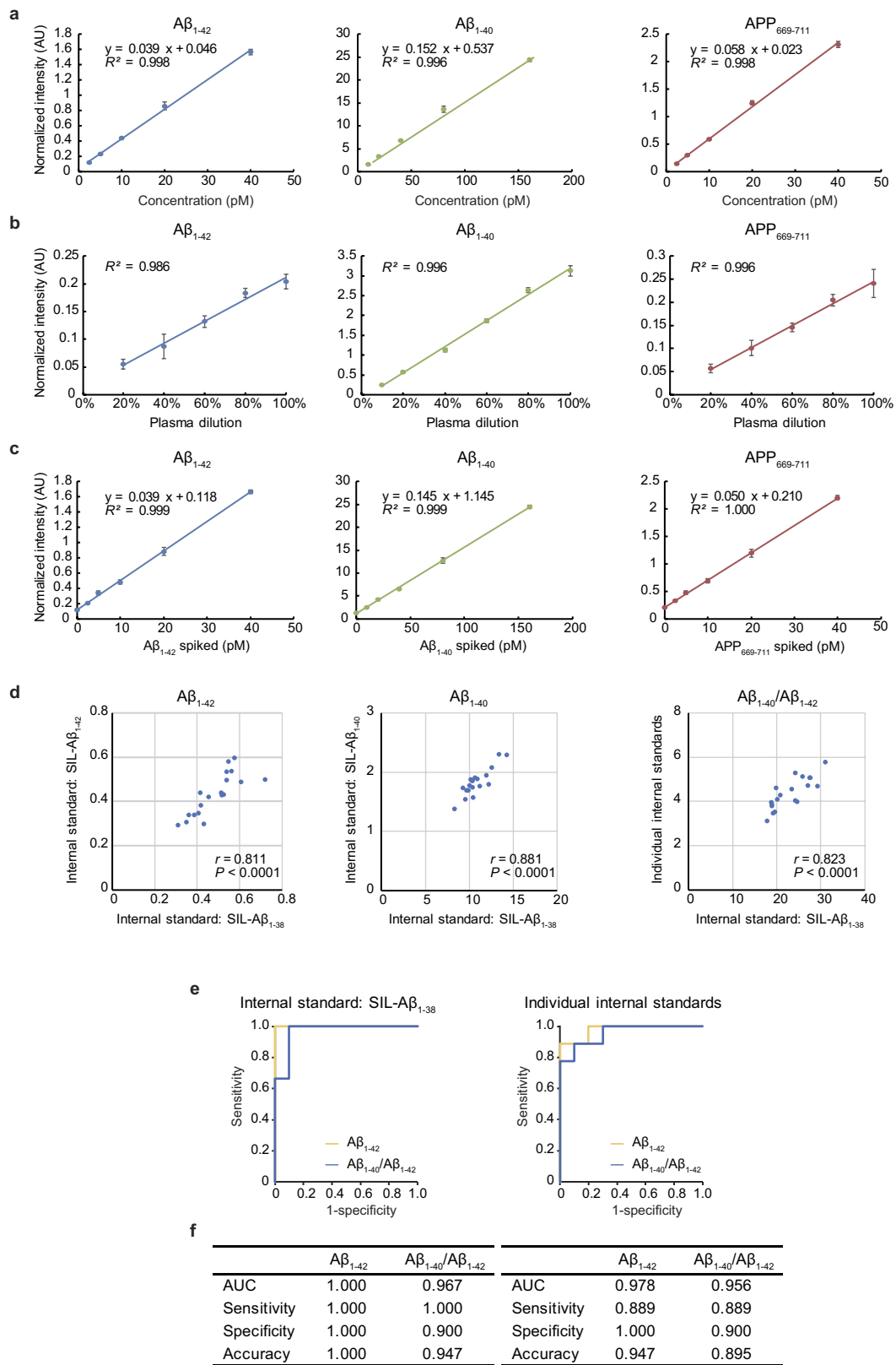
**Extended Data Figure 3 | Adjusted ROC analyses corresponding to Fig. 2. a,** ROC analyses for each biomarker when predicting individual  $A\beta^{+}/A\beta^{-}$  status for the discovery, validation, and combined data sets. Adjusted (age, gender, *APOE4*, and clinical category) analyses for the NCGG PIB discovery data (left), AIBL PIB validation data (middle), and AIBL overall (all tracers) validation data (right). See Extended Data Table 1a for detailed performance values. Data are from 121, 111 and 252 individuals for the NCGG PIB, AIBL PIB and AIBL overall data, respectively. **b,** Comparisons of biomarker performances within each analysis corresponding to the ROC curves in **a**. Each colour bar represents the AUC and 95% confidence interval. Statistically significant differences between two AUCs (DeLong test) and significant increments in predictive ability as assessed by NRI and IDI are indicated as in Fig. 2. All *P* values

are two-sided, and Bonferroni corrected (multiplied by the number of comparisons, 6). Note that the NRI in the comparison between  $A\beta_{1-40}/A\beta_{1-42}$  and the composite biomarker in NCGG data was negative (NRI =  $-0.382$ ) indicating that the reclassification ability is lower in the composite biomarker. **c,** Adjusted (age, gender, *APOE4* and clinical category) ROC analyses of the composite biomarker compared by different PET tracers; PIB (NCGG,  $n = 121$ , and AIBL,  $n = 111$ ), flutemetamol (AIBL,  $n = 81$ ), and florbetapir (AIBL,  $n = 60$ ). See Extended Data Table 1b for detailed performance values. **d, e,** Adjusted (age, gender, *APOE4*) ROC curves of the composite biomarker within the AD and MCI (**d**) and cognitively normal (**e**) groups. Sample sizes are the same as those listed in Fig. 2d, e; see Extended Data Table 1c for detailed performance values.



**Extended Data Figure 4 | Correlations between plasma biomarkers and brain  $A\beta$  burden: additional data related to Fig. 3a.** **a–c**, Biomarker values plotted against SUVR values from PIB-PET imaging;  $A\beta_{1-42}$  (**a**),  $APP_{669-711}/A\beta_{1-42}$  (**b**) and  $A\beta_{1-40}/A\beta_{1-42}$  (**c**). Data are from 121 (NCGG PIB, top) and 111 (AIBL PIB, bottom) individuals. Colours represent the clinical categories: AD, red; MCI, orange; cognitively normal, blue. The vertical dashed lines represent cut-off values of PIB-PET imaging (1.4),

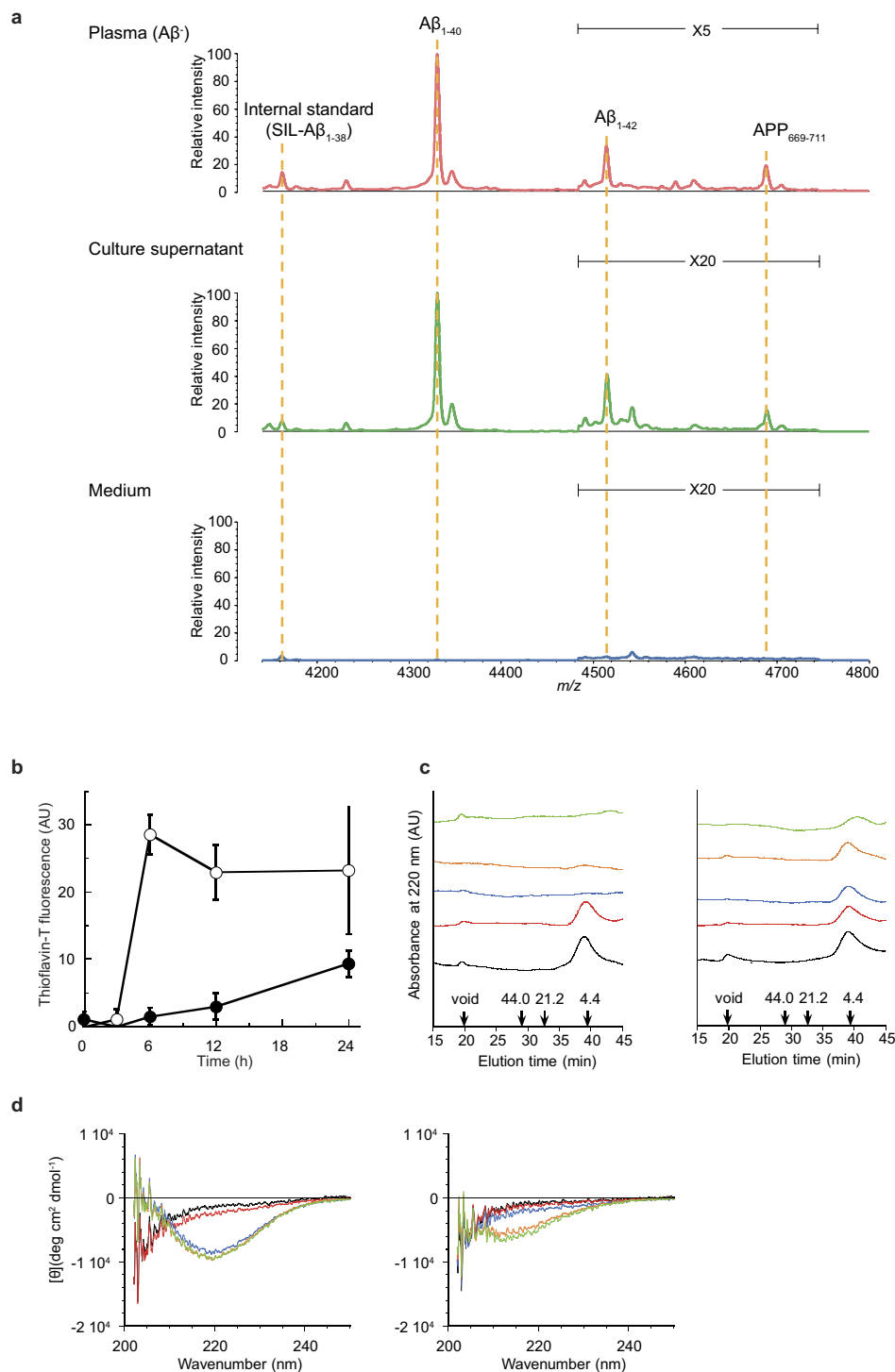
and horizontal dashed lines represent the common cut-off values of the plasma biomarkers estimated in Extended Data Fig. 7. **d**, A summary table for the correlation analyses. The sample sizes for each data set are: NCGG PIB,  $n = 121$ ; AIBL PIB,  $n = 111$ ; NCGG + AIBL PIB,  $n = 232$ ; AIBL FLUTE,  $n = 81$ ; AIBL FBP,  $n = 60$ ; and NCGG + AIBL overall,  $n = 373$ . Pearson's correlation coefficients ( $r$ ) and their significance (two-sided  $P$ ) are presented in the plots and the table.



Extended Data Figure 5 | See next page for caption.

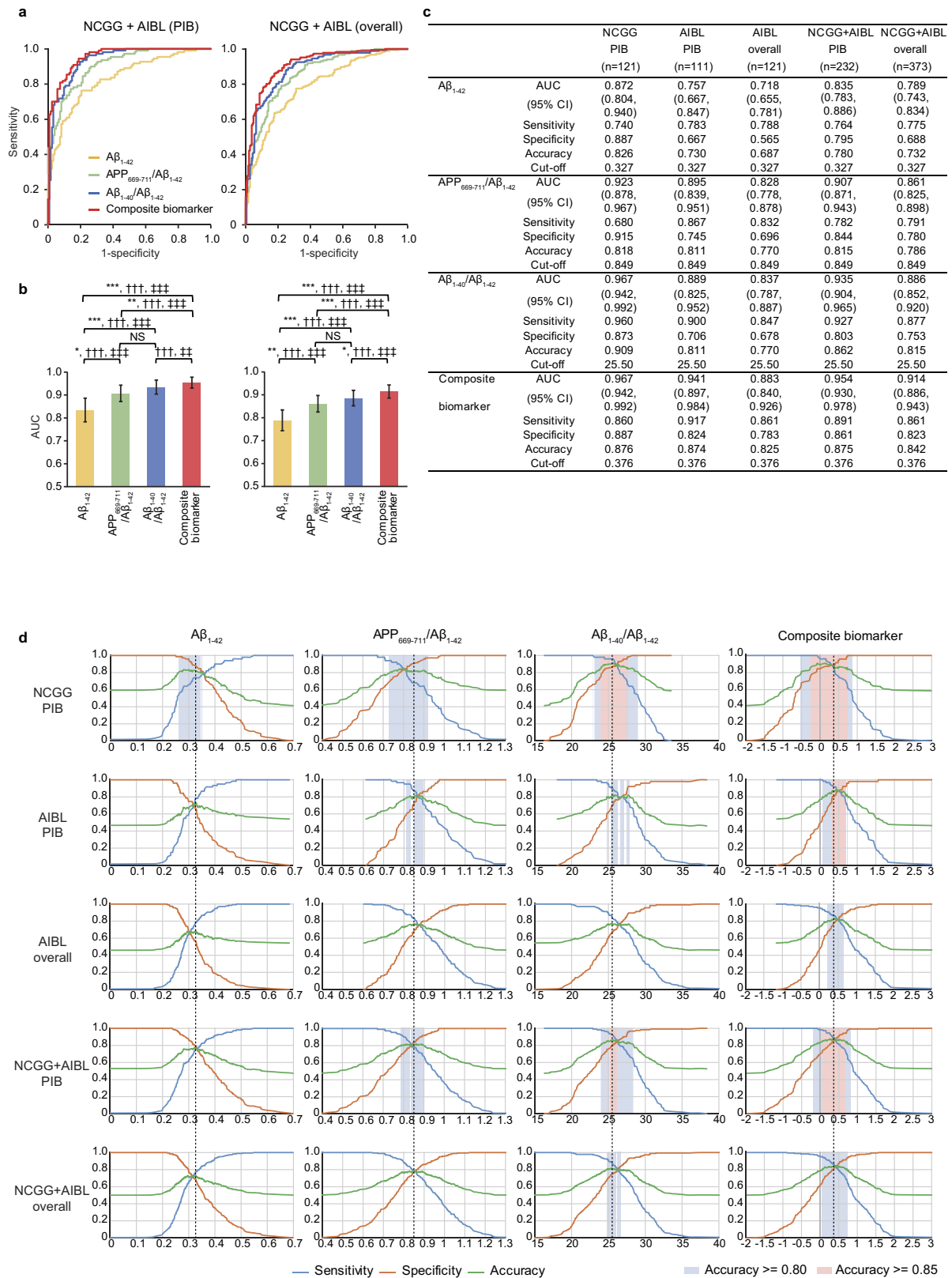
**Extended Data Figure 5 | Reliability of the IP-MS methods.** **a**, Standard curves of A $\beta_{1-42}$  (left), A $\beta_{1-40}$  (middle), and APP<sub>669-711</sub> (right) in PBS containing BSA. The standard curves were generated over a 2.5–40 pM range for A $\beta_{1-42}$  and APP<sub>669-711</sub>, and a 10–160 pM range for A $\beta_{1-40}$ . The linearity was evaluated with the coefficient of determination ( $R^2$ ). The error bars indicate the standard deviations of normalized intensities obtained from four mass spectra. The normalized intensities (AU) and signal-to-noise ratios at the lowest concentration were 0.119 AU and 10.9 for A $\beta_{1-42}$ , 0.152 AU and 16.1 for APP<sub>669-711</sub> and 1.56 AU and 165 for A $\beta_{1-40}$ , respectively. The lower limit of quantification referred to the lowest concentration at which A $\beta_{1-42}$ , APP<sub>669-711</sub> and A $\beta_{1-40}$  showed a signal-to-noise ratio greater than 10. Data reproducibility was confirmed by two additional experiments. **b**, Relationships between plasma dilution and normalized intensity of endogenous A $\beta_{1-42}$ , A $\beta_{1-40}$ , and APP<sub>669-711</sub>, which were contained in the human plasma. Normalized intensity indicates the mass spectrometry signal normalized with the signal for SIL-A $\beta_{1-38}$ . The linearity was evaluated with  $R^2$ . The error bars indicate the standard deviations of normalized intensities obtained from four mass spectra. The

data reproducibility was confirmed by one additional experiment. **c**, Signal linearity of plasma peptides spiked with synthetic A $\beta_{1-42}$ , synthetic A $\beta_{1-40}$ , and synthetic APP<sub>669-711</sub>. The plasma samples, which were spiked over a 2.5–40 pM range for A $\beta_{1-42}$  and APP<sub>669-711</sub>, and a 10–160 pM range for A $\beta_{1-40}$ , were prepared and measured by the IP-MS method. The linearity was evaluated with  $R^2$ . The error bars indicate the standard deviations of normalized intensities obtained from four mass spectra. The data reproducibility was confirmed by one additional experiment. **d**, Normalized signal intensity of A $\beta_{1-42}$ , A $\beta_{1-40}$ , and A $\beta_{1-40}$ /A $\beta_{1-42}$  in 19 subjects measured by two methods; using common internal standard SIL-A $\beta_{1-38}$  ( $x$  axis) and using corresponding SIL-peptides ( $y$  axis). Pearson's correlation coefficients ( $r$ ) and their significance (two-sided  $P$ ) are presented in the plots. The experiments were performed once. **e**, ROC analyses for A $\beta_{1-42}$  and A $\beta_{1-40}$ /A $\beta_{1-42}$  to distinguish between A $\beta^+$  and A $\beta^-$  individuals ( $n = 19$ ) of the two methods; using the common internal standard SIL-A $\beta_{1-38}$  (left) and using the corresponding SIL-peptides (right). **f**, Tables showing the performance values corresponding to **e**, as determined by ROC analyses and Youden's index.



**Extended Data Figure 6 | Cellular and molecular characteristics of APP<sub>669-711</sub>.** **a**, Results of additional experiment 1 (Supplementary Information).  $A\beta$ -related peptides produced from human neuroblastoma cell line. MALDI-TOF mass spectra of  $A\beta$ -related peptides in human plasma (top), BE(2)-C cell culture supernatant (middle) and medium without cell culture (bottom). Representative spectra from five experiments are shown. The theoretical  $m/z$  values of peptides are 4,330.9 for  $A\beta_{1-40}$ , 4,515.1 for  $A\beta_{1-42}$ , and 4,689.4 for APP<sub>669-711</sub>. SIL- $A\beta_{1-38}$  was used as an internal standard for the normalization of mass spectra. **b–d**, Results of additional experiment 2 (Supplementary Information). **b**, Kinetics of fibril formation.  $A\beta_{1-42}$  (15  $\mu$ M, open circles) or APP<sub>669-711</sub> (15  $\mu$ M, closed circles) were incubated in PBS at 37 °C. Fibril formation

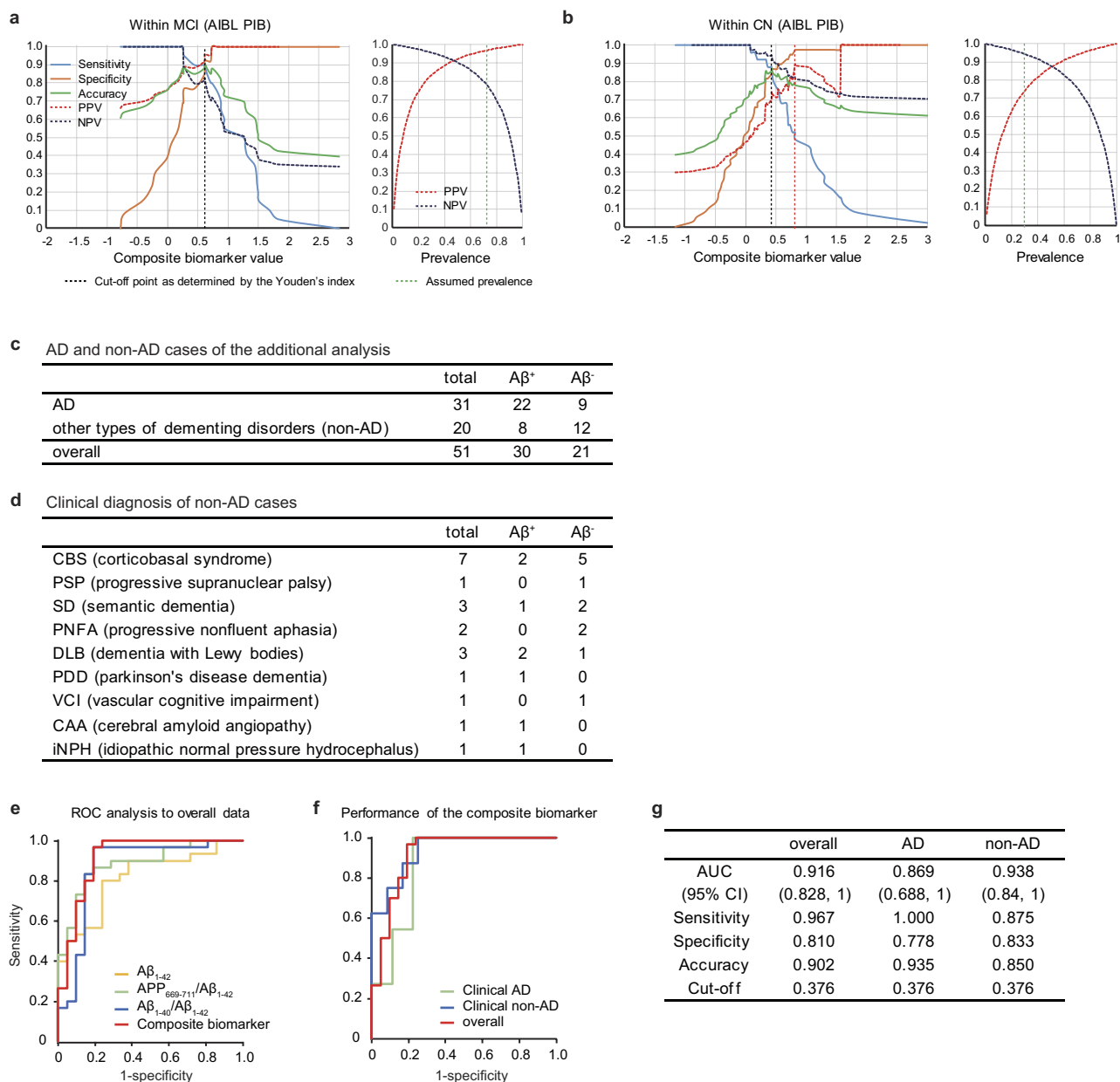
was monitored using the thioflavin T spectroscopic assay. Data are mean  $\pm$  s.d. from four ( $A\beta_{1-42}$ ) or five (APP<sub>669-711</sub>) experiments. **c**, Size exclusion chromatography.  $A\beta_{1-42}$  (15  $\mu$ M, left) or APP<sub>669-711</sub> (15  $\mu$ M, right) were incubated in PBS at 37 °C, and the supernatants were centrifuged (10,000g for 10 min) and subjected to size-exclusion chromatography (Sephacryl S-300 HR) at 0 (black), 3 (red), 6 (blue), 12 (orange), and 24 h (green). The elution times for molecular mass standards (kDa) are indicated by arrows. **d**, Changes in secondary structure during peptide aggregation.  $A\beta_{1-42}$  (15  $\mu$ M, left) and APP<sub>669-711</sub> (15  $\mu$ M, right) were incubated in PBS at 37 °C, and circular dichroism spectra were measured at 0 (black), 3 (red), 6 (blue), 12 (orange), and 24 h (green). Experiments in **b–d** were each performed once.



Extended Data Figure 7 | See next page for caption.

**Extended Data Figure 7 | Common cut-off values are applicable for both NCGG and AIBL data sets.** **a**, Unadjusted ROC analyses for each biomarker when predicting individual  $A\beta^+/A\beta^-$  status for the NCGG + AIBL PIB (left,  $n = 232$ ) and NCGG + AIBL overall (right,  $n = 373$ ) data sets. **b**, Comparisons of biomarker performances within each analysis corresponding to the ROC curves in **a**. Each colour bar represents the AUC and 95% confidence. Statistically significant differences between two AUCs (DeLong test) and significant increments in predictive ability as assessed by NRI and IDI are indicated as in Fig. 2. All  $P$  values are two-sided, and Bonferroni corrected (multiplied by the number of comparisons, 6). **c**, Biomarker performances when applying the same cut-off values to each data set. For each biomarker, an optimal

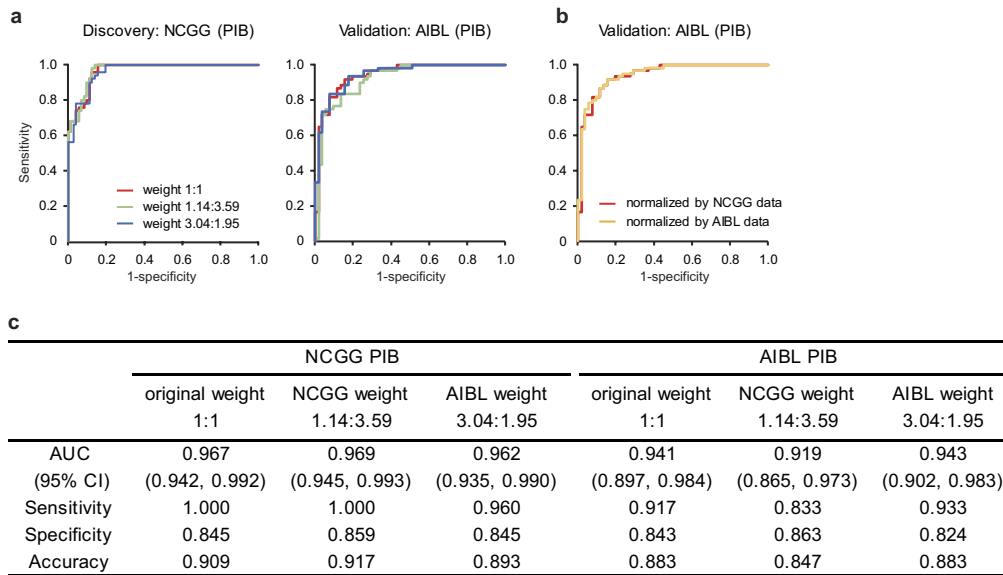
common cut-off value was determined by the Youden's index of the ROC analysis for the NCGG + AIBL overall data set. The sensitivity, specificity and accuracy were then calculated at the common cut-off point for each biomarker in all data sets. **d**, Diagnostic performance plots for  $A\beta_{1-42}$ ,  $APP_{669-711}/A\beta_{1-42}$ ,  $A\beta_{1-40}/A\beta_{1-42}$  and the composite biomarker. Each row from top to bottom shows the plots for the NCGG PIB, AIBL PIB, AIBL overall, NCGG + AIBL PIB, and NCGG + AIBL overall data (unadjusted), respectively. Sensitivity (orange), specificity (blue) and accuracy (green) were plotted using the values of the biomarkers ( $x$  axis). The vertical dashed lines indicate the common cut-off values as shown in **c**. The blue and pink shaded squares indicate ranges in which a biomarker can perform with at least 80% and 85% accuracy, respectively.



### Extended Data Figure 8 | Possible clinical utility of the plasma biomarker.

**a, b**, Diagnostic performance plots for the composite biomarker in two specific settings (see Supplementary Discussion). **a**, Diagnostic performance plots for subjects with MCI in the AIBL PIB unadjusted data ( $n = 33$ ) (left). The prevalence of A $\beta$ -positivity for subjects with MCI was assumed to be 66%. Sensitivity (orange), specificity (blue), accuracy (green), PPV (red, dashed) and NPV (dark blue, dashed) were plotted against the values of the composite biomarker ( $x$  axis). The black vertical dashed line indicates a cut-off point as determined by the Youden's index ( $y$  point) in the AIBL PIB data. At the  $y$  point, the sensitivity and specificity were 0.900 and 0.923, respectively. With these values, relationships between the prevalence and PPV or NPV were plotted (right). Note that these data do not correspond to the ROC analysis shown in Fig. 2d, because this diagnostic performance plot analysis does not contain subjects with AD. **b**, Diagnostic performance

plots for cognitively normal subjects in the AIBL PIB unadjusted data ( $n = 63$ ) (left). The prevalence of A $\beta$ -positivity in general elderly people was assumed to be 30%. At the  $y$  point, the sensitivity and specificity were 0.880 and 0.868, respectively. With these values, relationships between the prevalence and PPV or NPV were also plotted (right). **c–g**, Results of the additional analysis for subjects with and without AD (see Supplementary Discussion). **c**, Sample numbers for subjects with and without AD. **d**, Detailed clinical diagnoses of subjects without AD. **e**, ROC analyses for each plasma biomarker in the overall data ( $n = 51$ ). **f**, ROC analysis of the composite biomarker in the overall ( $n = 51$ ), AD ( $n = 31$ ), and non-AD ( $n = 20$ ) groups. **g**, Performance of the composite biomarker using the common cut-off value. The AUC values were computed from the ROC analysis and sensitivity, specificity, and accuracy were computed by applying the common cut-off value for the composite biomarker (0.376).



**Extended Data Figure 9 | Optimal generation of the composite biomarker.** Unadjusted ROC analyses of the composite biomarkers generated by different weightings (see Supplementary Discussion). **a**, Comparisons of the composite biomarkers generated by different weights for APP<sub>669-711</sub>/A $\beta$ <sub>1-42</sub> and A $\beta$ <sub>1-40</sub>/A $\beta$ <sub>1-42</sub> normalized values (z-scores) in the discovery NCGG PIB data ( $n = 121$ , left) and validation AIBL PIB data ( $n = 111$ , right). The composite biomarker generated by the original weight (1:1) is coloured red, the weight estimated by the

NCGG data (1.14:3.59) green, and the weight estimated by the AIBL data (3.04:1.95) blue. **b**, A comparison of the composite biomarkers generated by using different reference databases. The original composite biomarker (normalized by NCGG data) is coloured red, and the alternative composite biomarker normalized by AIBL data is orange. **c**, Summary of the ROC analyses. The AUCs and the representative best values of sensitivity, specificity and accuracy for these biomarkers as determined by the Youden's index are shown.

**Extended Data Table 1 | Performance values of the plasma biomarkers**

**a** Performances of each biomarker as analyzed by the ROC analysis

|  |             | Unadjusted     |                |                 | Adjusted       |                |                 |
|--|-------------|----------------|----------------|-----------------|----------------|----------------|-----------------|
|  |             | Discovery      |                | Validation      | Discovery      |                | Validation      |
|  |             | NCGG<br>PIB    | AIBL<br>PIB    | AIBL<br>overall | NCGG<br>PIB    | AIBL<br>PIB    | AIBL<br>overall |
| Aβ <sub>1-42</sub>                         | AUC         | 0.872          | 0.757          | 0.718           | 0.913          | 0.812          | 0.797           |
|  | (95% CI)    | (0.804, 0.94)  | (0.667, 0.847) | (0.655, 0.781)  | (0.853, 0.973) | (0.733, 0.891) | (0.743, 0.851)  |
|  | Sensitivity | 0.740          | 0.783          | 0.635           | 0.820          | 0.800          | 0.774           |
|  | Specificity | 0.887          | 0.667          | 0.739           | 0.958          | 0.706          | 0.713           |
|  | Accuracy    | 0.826          | 0.730          | 0.683           | 0.901          | 0.757          | 0.746           |
|  | Cut-off     | 0.324          | 0.328          | 0.295           | 0.533          | 0.512          | 0.517           |
| APP <sub>669-711</sub> /Aβ <sub>1-42</sub> | AUC         | 0.923          | 0.895          | 0.828           | 0.933          | 0.905          | 0.854           |
|  | (95% CI)    | (0.878, 0.967) | (0.839, 0.951) | (0.778, 0.878)  | (0.893, 0.973) | (0.852, 0.958) | (0.808, 0.900)  |
|  | Sensitivity | 0.900          | 0.850          | 0.832           | 0.760          | 0.900          | 0.766           |
|  | Specificity | 0.803          | 0.784          | 0.696           | 0.944          | 0.765          | 0.817           |
|  | Accuracy    | 0.843          | 0.820          | 0.770           | 0.868          | 0.838          | 0.790           |
|  | Cut-off     | 0.779          | 0.861          | 0.849           | 0.651          | 0.443          | 0.571           |
| Aβ <sub>1-40</sub> /Aβ <sub>1-42</sub>     | AUC         | 0.967          | 0.889          | 0.837           | 0.979          | 0.897          | 0.851           |
|  | (95% CI)    | (0.942, 0.992) | (0.825, 0.952) | (0.787, 0.887)  | (0.961, 0.997) | (0.837, 0.958) | (0.804, 0.898)  |
|  | Sensitivity | 0.960          | 0.733          | 0.657           | 0.900          | 0.833          | 0.745           |
|  | Specificity | 0.873          | 0.922          | 0.896           | 0.944          | 0.843          | 0.826           |
|  | Accuracy    | 0.909          | 0.820          | 0.766           | 0.926          | 0.838          | 0.782           |
|  | Cut-off     | 25.469         | 27.656         | 27.723          | 0.550          | 0.528          | 0.560           |
| Composite biomarker                        | AUC         | 0.967          | 0.941          | 0.883           | 0.974          | 0.940          | 0.888           |
|  | (95% CI)    | (0.942, 0.992) | (0.897, 0.984) | (0.840, 0.926)  | (0.953, 0.995) | (0.898, 0.982) | (0.847, 0.929)  |
|  | Sensitivity | 1.000          | 0.917          | 0.854           | 0.920          | 0.833          | 0.876           |
|  | Specificity | 0.845          | 0.843          | 0.800           | 0.915          | 0.922          | 0.774           |
|  | Accuracy    | 0.909          | 0.883          | 0.829           | 0.917          | 0.874          | 0.829           |
|  | Cut-off     | -0.079         | 0.425          | 0.425           | 0.407          | 0.663          | 0.466           |

**b** Performance of the composite biomarker for each PET tracer

|             | Unadjusted     |                |                |                | Adjusted       |                |                |                |
|-------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|             | Discovery      | Validation     |                |                | Discovery      | Validation     |                |                |
|             | NCGG PIB       | AIBL PIB       | AIBL FLUTE     | AIBL FBP       | NCGG PIB       | AIBL PIB       | AIBL FLUTE     | AIBL FBP       |
| AUC         | 0.967          | 0.941          | 0.829          | 0.864          | 0.974          | 0.940          | 0.849          | 0.850          |
| (95% CI)    | (0.942, 0.992) | (0.897, 0.984) | (0.737, 0.921) | (0.772, 0.957) | (0.953, 0.995) | (0.898, 0.982) | (0.764, 0.934) | (0.750, 0.950) |
| Sensitivity | 1.000          | 0.917          | 0.787          | 1.000          | 0.920          | 0.833          | 0.809          | 1.000          |
| Specificity | 0.845          | 0.843          | 0.824          | 0.633          | 0.915          | 0.922          | 0.794          | 0.700          |
| Accuracy    | 0.909          | 0.883          | 0.802          | 0.817          | 0.917          | 0.874          | 0.802          | 0.850          |
| Cut-off     | -0.079         | 0.425          | 0.491          | 0.128          | 0.407          | 0.663          | 0.570          | 0.290          |

**c** Performance of the composite biomarker within each clinical category

| Unadjusted        |                |                |                            | Adjusted          |                |                            |
|-------------------|----------------|----------------|----------------------------|-------------------|----------------|----------------------------|
| Within AD and MCI |                |                |                            | Within AD and MCI |                |                            |
|                   | PIB NCGG       | PIB AIBL       | <sup>18</sup> F Aβ tracers | PIB NCGG          | PIB AIBL       | <sup>18</sup> F Aβ tracers |
| AUC               | 0.980          | 0.974          | 0.894                      | 0.983             | 0.978          | 0.905                      |
| (95% CI)          | (0.953, 1.000) | (0.937, 1.000) | (0.799, 0.989)             | (0.959, 1.000)    | (0.945, 1.000) | (0.818, 0.992)             |
| Sensitivity       | 0.975          | 0.914          | 0.944                      | 0.875             | 0.886          | 0.944                      |
| Specificity       | 0.895          | 0.923          | 0.750                      | 1.000             | 1.000          | 0.750                      |
| Accuracy          | 0.949          | 0.917          | 0.896                      | 0.915             | 0.917          | 0.896                      |
| Cut-off f         | 0.192          | 0.610          | 0.124                      | 0.868             | 0.863          | 0.549                      |
| Within CN         |                |                |                            | Within CN         |                |                            |
|                   | PIB NCGG       | PIB AIBL       | <sup>18</sup> F Aβ tracers | PIB NCGG          | PIB AIBL       | <sup>18</sup> F Aβ tracers |
| AUC               | 0.912          | 0.917          | 0.800                      | 0.942             | 0.873          | 0.786                      |
| (95% CI)          | (0.840, 0.984) | (0.849, 0.985) | (0.705, 0.895)             | (0.886, 0.999)    | (0.787, 0.958) | (0.692, 0.879)             |
| Sensitivity       | 1.000          | 0.880          | 0.780                      | 1.000             | 0.840          | 0.780                      |
| Specificity       | 0.865          | 0.868          | 0.808                      | 0.827             | 0.816          | 0.731                      |
| Accuracy          | 0.887          | 0.873          | 0.796                      | 0.855             | 0.825          | 0.753                      |
| Cut-off f         | -0.112         | 0.425          | 0.491                      | 0.130             | 0.355          | 0.402                      |

**a**, Performances of each biomarker as analysed by the ROC analyses corresponding to Fig. 2a (unadjusted analysis, left) and Extended Data Fig. 3a (adjusted analysis, right). All values except AUC are the representative best values for each ROC analysis at a cut-off point determined using Youden's index (see Methods). The cut-off values for the adjusted analyses are predictive values of the logistic regression analyses. **b**, Performance of the composite biomarker for each PET tracer. The left and right panels correspond to the results of Fig. 2c (unadjusted) and Extended Data Fig. 3c (adjusted), respectively. **c**, The performance of the composite biomarker within each clinical category. The left panel (unadjusted) corresponds to the results of Fig. 2d (top) and Fig. 2e (bottom). The right panel (adjusted) corresponds to the results of Extended Data Fig. 3d (top) and Extended Data Fig. 3e (bottom).

# Innate and adaptive lymphocytes sequentially shape the gut microbiota and lipid metabolism

Kairui Mao<sup>1</sup>, Antonio P. Baptista<sup>1</sup>, Samira Tamoutounour<sup>2</sup>, Lenan Zhuang<sup>3</sup>, Nicolas Bouladoux<sup>2</sup>, Andrew J. Martins<sup>4</sup>, Yuefeng Huang<sup>5</sup>, Michael Y. Gerner<sup>6</sup>, Yasmine Belkaid<sup>2</sup> & Ronald N. Germain<sup>1</sup>

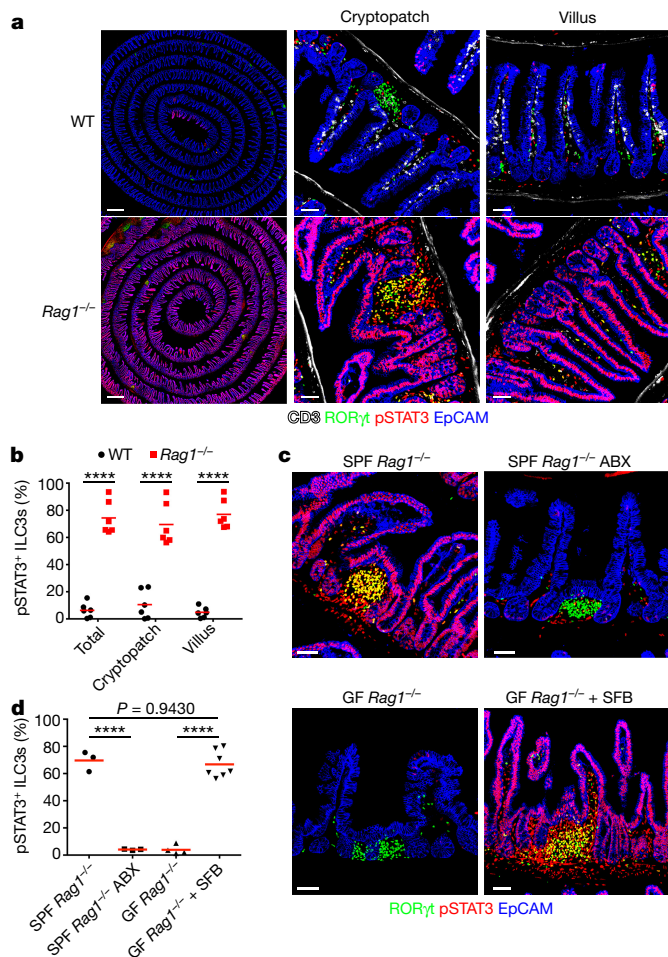
**The mammalian gut is colonized by numerous microorganisms collectively termed the microbiota, which have a mutually beneficial relationship with their host<sup>1–3</sup>. Normally, the gut microbiota matures during ontogeny to a state of balanced commensalism marked by the absence of adverse inflammation<sup>4,5</sup>. Subsets of innate lymphoid cells (ILCs) and conventional T cells are considered to have redundant functions in containment and clearance of microbial pathogens<sup>6,7</sup>, but how these two major lymphoid-cell populations each contribute to shaping the mature commensal microbiome and help to maintain tissue homeostasis has not been determined. Here we identify, using advanced multiplex quantitative imaging methods, an extensive and persistent phosphorylated-STAT3 signature in group 3 ILCs and intestinal epithelial cells that is induced by interleukin (IL)-23 and IL-22 in mice that lack CD4<sup>+</sup> T cells. By contrast, in immune-competent mice, phosphorylated-STAT3 activation is induced only transiently by microbial colonization at weaning. This early signature is extinguished as CD4<sup>+</sup> T cell immunity develops in response to the expanding commensal burden. Physiologically, the persistent IL-22 production from group 3 ILCs that occurs in the absence of adaptive CD4<sup>+</sup> T-cell activity results in impaired host lipid metabolism by decreasing lipid transporter expression in the small bowel. These findings provide new insights into how innate and adaptive lymphocytes operate sequentially and in distinct ways during normal development to establish steady-state commensalism and tissue metabolic homeostasis.**

To study the state and activity of innate and adaptive immune cells *in situ* in the context of the microbiota, we used quantitative multiplex immunohistochemistry (histo-cytometry)<sup>8–10</sup>. Staining for both cell phenotype and molecules that indicate active cytokine signalling enabled us to simultaneously examine the cell distribution, cytokine production and topography of cytokine responses in a complex tissue. Because many relevant cytokines induce phosphorylation of STAT3<sup>11</sup>, we focused on the presence and distribution of phosphorylated STAT3 (pSTAT3) in the distal small intestine of animals with an intact (wild type) or adaptive-lymphocyte-deficient (*Rag1*<sup>−/−</sup>) immune system under specific pathogen-free conditions. Using this approach, we observed substantial pSTAT3 in both CD3<sup>−</sup>RORγt<sup>+</sup>-group 3 innate lymphoid cells (ILC3s) and nearly all EpCAM<sup>+</sup> intestinal epithelial cells (IECs) in the distal small intestine of *Rag1*<sup>−/−</sup> but not wild-type mice (Fig. 1a). RORγt<sup>+</sup> ILC3s occur as a number of subpopulations with different locations in the gut: CCR6<sup>+</sup> lymphoid-tissue inducer-like ILC3s reside mainly in cryptopatches and isolated lymphoid follicles, while NKp46<sup>+</sup> ILC3s and NKp46<sup>−</sup> ILC3s are mostly located in the lamina propria<sup>12</sup>. ILC3s in both locations were equally activated in *Rag1*<sup>−/−</sup> mice (Fig. 1a, b and Extended Data Fig. 1). Because the small intestine is colonized with commensal microbiota, we examined

whether these microorganisms had a role in the pSTAT3 signature by treating specific pathogen-free *Rag1*<sup>−/−</sup> mice with broad-spectrum antibiotics or using germ-free *Rag1*<sup>−/−</sup> mice. All pSTAT3 signals in ILC3s and IECs were eliminated in these mice (Fig. 1c, d). Segmented filamentous bacteria (SFB) attach directly to IECs in the distal small intestine and contribute to TH17 cell differentiation and ILC3 activation *in vivo*<sup>13,14</sup>. We therefore mono-colonized germ-free *Rag1*<sup>−/−</sup> mice with SFB, which resulted in robust STAT3 activation in ILC3s and IECs (Fig. 1c, d). We conclude that STAT3 phosphorylation in ILC3s and IECs arises from signalling induced by microbiota, and that defined microbes such as SFB have a major role in this signalling in *Rag1*<sup>−/−</sup> mice.

IL-23 and IL-22 are functionally linked cytokines with the potential to induce pSTAT3 in ILC3s and IECs, respectively<sup>15,16</sup>. We therefore examined pSTAT3 in *Rag1*<sup>−/−</sup> mice that were also deficient in *Il23a* (which codes for the IL-23p19 subunit) or *Il22*. In the absence of IL-23, pSTAT3 was no longer detected in ILC3s or IECs. By contrast, *Il22*<sup>−/−</sup>*Rag1*<sup>−/−</sup> mice had a similar proportion of pSTAT3<sup>+</sup> ILC3s as *Rag1*<sup>−/−</sup> mice, but lacked pSTAT3<sup>+</sup> IECs. The pSTAT3 signature in ILC3s and IECs was independent of IL-6 (Extended Data Fig. 2a, b). These data reveal that pSTAT3 generation in ILC3s depends on IL-23 but not on IL-6 or IL-22, whereas pSTAT3 in IECs depends on both IL-23 and IL-22. Because ILC3s and not IECs express the IL-23 receptor, it seemed likely that STAT3 activation in ILC3s and IECs were sequential events downstream of IL-23 and IL-22, respectively. To test this hypothesis, we examined *Rorc*(γt)<sup>GFP/GFP</sup>*Rag1*<sup>−/−</sup> mice, which specifically lack ILC3s. The absence of pSTAT3 in IECs of these mice suggests that ILC3s are critical for STAT3 activation in epithelial cells in the small intestine of *Rag1*<sup>−/−</sup> mice (Extended Data Fig. 2c). Examination of *Rag1*<sup>−/−</sup> mice expressing an *Il22*-tdTomato reporter showed that pSTAT3<sup>+</sup> ILC3s produced IL-22 (Extended Data Fig. 2d). To investigate the source of IL-23, we isolated different myeloid-cell populations from the small intestine of wild-type and *Rag1*<sup>−/−</sup> mice<sup>17</sup> (Extended Data Fig. 3a). CD11b<sup>+</sup> conventional dendritic cells, as well as CCR2<sup>+</sup> monocytes and monocyte-derived dendritic cells, can all express *Il23a*, but CCR2<sup>+</sup> cells from *Rag1*<sup>−/−</sup> mice had higher *Il23a* expression than those from wild-type mice, whereas there was no significant difference in *Il23a* expression in the CD11b<sup>+</sup> populations (Extended Data Fig. 3b). To determine whether IL-23 production from CCR2<sup>+</sup> myeloid cells was responsible for the pSTAT3 signature in *Rag1*<sup>−/−</sup> mice, we depleted these cells by injecting these mice with a CCR2 antibody. After two weeks, there were no pSTAT3<sup>+</sup> ILC3s or IECs in antibody-treated *Rag1*<sup>−/−</sup> mice (Extended Data Fig. 3c), which indicates that IL-23 production in CCR2<sup>+</sup> cells is critical for STAT3 activation in the small intestine of *Rag1*<sup>−/−</sup> mice. Similar results were obtained using a more broadly depleting antibody against GR1 (Extended Data Fig. 3c).

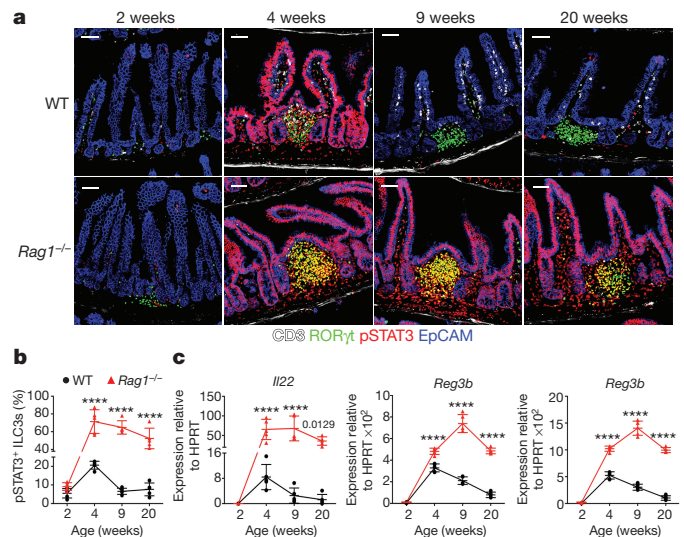
<sup>1</sup>Lymphocyte Biology Section, Laboratory of Systems Biology, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Mucosal Immunology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>3</sup>Laboratory of Endocrinology and Receptor Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>4</sup>Systems Genomics and Bioinformatics Unit, Laboratory of Systems Biology, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>5</sup>Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>6</sup>Department of Immunology, University of Washington School of Medicine, Seattle, Washington 98109, USA.



**Figure 1 | pSTAT3<sup>+</sup> ILC3s and IECs induced by microbiota in small intestine of *Rag1*<sup>-/-</sup> mice.** **a–d**, Immunofluorescence staining of ileum (**a**, **c**) and quantification of pSTAT3<sup>+</sup> ILC3s (**b**, **d**) from wild-type (WT) or *Rag1*<sup>-/-</sup> mice ( $n = 6$ ; **a**, **b**), and from specific pathogen-free (SPF) *Rag1*<sup>-/-</sup> mice with or without antibiotics treatment ( $n = 3$ ), germ-free *Rag1*<sup>-/-</sup> mice ( $n = 4$ ) and germ-free *Rag1*<sup>-/-</sup> mice mono-colonized with SFB ( $n = 7$ ; **c**, **d**). ABX, antibiotics; GF, germ-free. Results are representative of two (**c**, **d**) or three (**a**, **b**) independent experiments. Bars indicate mean; \*\*\*\* $P < 0.0001$ , otherwise exact  $P$  values are shown; two-tailed Student's  $t$ -test (**b**) or one-way ANOVA (**d**). Scale bars, 50  $\mu\text{m}$  (except **a**, left, 500  $\mu\text{m}$ ).

Together, IL-23 and IL-22 constitute a circuit that involves ILC3s and produces strong signalling in IECs in animals that lack an adaptive immune system. Though such a circuit was previously reported upon pathogen infection of *Rag1*-deficient mice<sup>15</sup>, these new findings show that ILC3s and IECs are robustly and persistently activated by the commensal microbiota in the absence of adaptive immunity.

Responses to microbes typically involve sequential activity of the innate and then adaptive immune systems. However, whether ILC responses that are thought to qualitatively parallel those of CD4<sup>+</sup> T cells with which they share master transcription factor expression (for example, RORγt expression) operate sequentially or in parallel is unknown. The difference in IL-22-dependent IEC signalling between wild-type and *Rag1*<sup>-/-</sup> adult mice suggested that in wild-type mice, innate cells might be activated by the microbiota before an effective adaptive immune response develops. To explore this possibility, we examined ileal pSTAT3 in wild-type and *Rag1*<sup>-/-</sup> progeny of SFB<sup>+</sup> mothers. In neonatal mice, which have a less diverse microbiota and limited SFB colonization<sup>18</sup>, neither wild-type nor *Rag1*<sup>-/-</sup> mice had pSTAT3<sup>+</sup> ILC3s or IECs. However, shortly after weaning, when bacterial colonization and expansion of the SFB population occurs, there was substantial activation of ILC3s and STAT3 signalling in

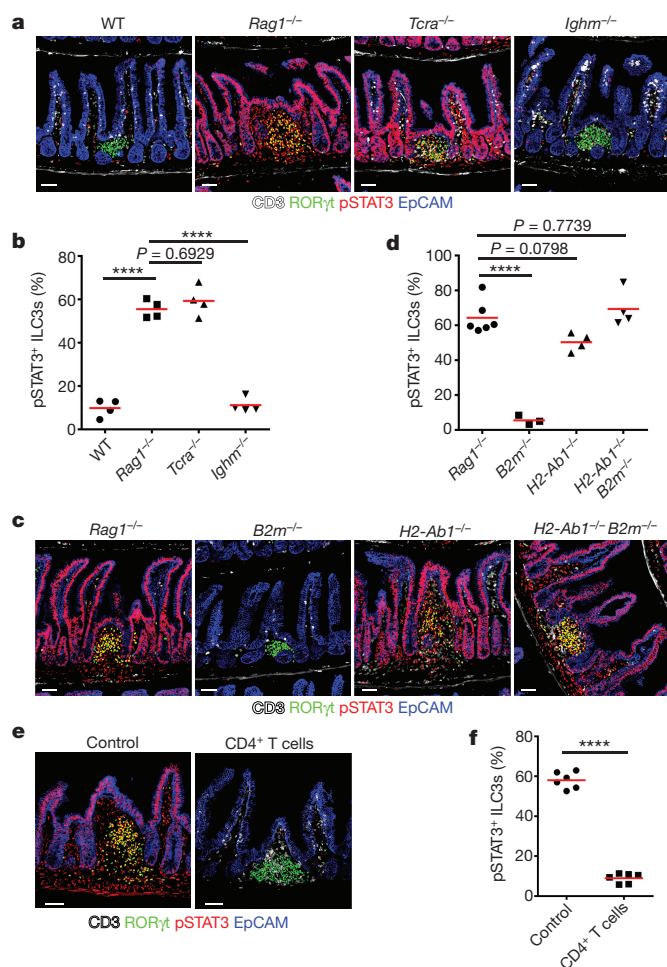


**Figure 2 | Transient activation of ILC3s and IECs shortly after weaning in wild-type mice.** **a**, Immunofluorescence staining of ileum from wild-type and *Rag1*<sup>-/-</sup> mice at the indicated ages ( $n = 5$ ). **b**, Quantification of pSTAT3<sup>+</sup> ILC3s in **a**. **c**, Expression of the indicated genes in the ileum from mice as in **a** ( $n = 5$ ). Results are representative of two independent experiments. Mean  $\pm$  s.d.; \*\*\*\* $P < 0.0001$ , otherwise exact  $P$  values are shown; two-way ANOVA. Scale bars, 50  $\mu\text{m}$ .

IECs in wild-type mice, although this was less robust than in *Rag1*<sup>-/-</sup> mice (Fig. 2a, b). As the adaptive immune system of wild-type mice matured, ILC3s were no longer activated, whereas the activation persisted at high levels in ILC3s in *Rag1*<sup>-/-</sup> mice (Fig. 2a, b). Expression of *Il22* and host-defence genes correlated precisely with ILC3 activation (Fig. 2c). These data indicate that during ontogeny, innate lymphoid cells operate before the adaptive system has fully developed<sup>19,20</sup>. The emerging adaptive response then largely silences the ILC response and establishes a homeostatic state of non-inflammatory commensalism. Notably, the absence of the pSTAT3 signature in IECs from adult wild-type mice indicates that the effector mechanisms of the adaptive immune response to commensals is mechanistically distinct from that of ILC3s and that the existing paradigm that equates the effector functions of ILCs with subsets of CD4<sup>+</sup>-effector-T cells is not universally applicable<sup>21,22</sup>.

To study the potentially distinct effects of innate and adaptive lymphocytes on the microbiota, we measured the abundance of different bacterial species in the small intestines of co-housed wild-type, *Rag1*<sup>-/-</sup> and *Il23a*<sup>-/-</sup> *Rag1*<sup>-/-</sup> mice. In the absence of adaptive immunity, most bacteria, including SFB, were increased in *Rag1*<sup>-/-</sup> mice. However, in the absence of STAT3 activation, SFB abundance was further increased in *Il23a*<sup>-/-</sup> *Rag1*<sup>-/-</sup> mice (Extended Data Fig. 4a). Although both innate and adaptive lymphocytes control the quantity of SFB, scanning electron microscopy revealed that SFB have markedly different morphology depending on whether adaptive lymphocytes are present. In the ileum of *Rag1*<sup>-/-</sup> mice, activated ILC3s prevented development of SFB into long filamentous forms, and the adaptive lymphocytes limited the number of SFB that attached to epithelial cells (Extended Data Fig. 4b, c). These findings show that innate and adaptive lymphocytes adopt different strategies to regulate the commensal state, with the activity of adaptive lymphocytes dominating over that of innate lymphoid cells under these conditions.

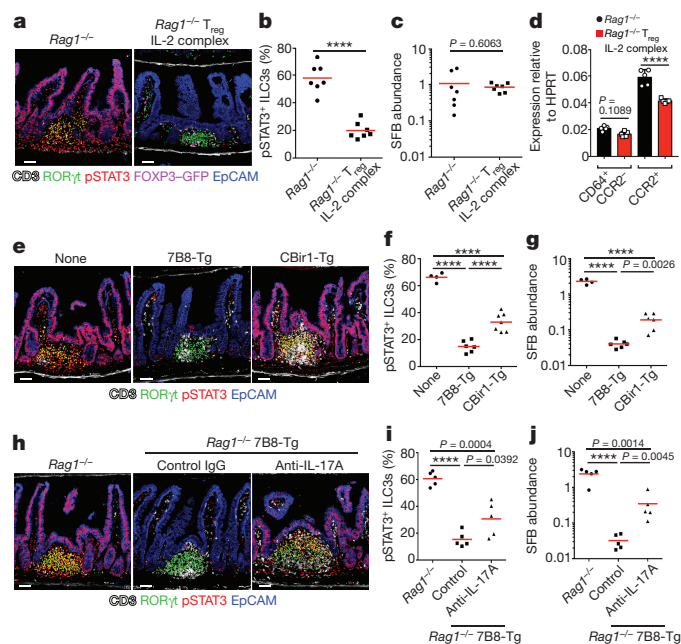
We next investigated which host adaptive immune cells prevent ILC3 activation and epithelial-cell signalling. First, we compared the distribution of pSTAT3 in the ileum of wild-type, *Rag1*<sup>-/-</sup>, *Tcra*<sup>-/-</sup> (T cell-deficient) and *Ighm*<sup>-/-</sup> (B cell-deficient) mice. Neither T cell nor B cell-deficient mice had pSTAT3<sup>+</sup> ILC3s or IECs (Extended Data Fig. 5a, b). As the pSTAT3 signature in *Rag1*<sup>-/-</sup> mice depended



**Figure 3 | Role of CD4<sup>+</sup> T cells in controlling ILC3 and IEC activation.** Immunofluorescence staining of ileum (a, c and e) and quantification of pSTAT3<sup>+</sup> ILC3s (b, d and f) from co-housed wild-type, *Rag1*<sup>-/-</sup>, *Tcr*<sup>-/-</sup> and *Ighm*<sup>-/-</sup> mice ( $n = 4$ ; a, b); co-housed *Rag1*<sup>-/-</sup> ( $n = 6$ ), *B2m*<sup>-/-</sup> ( $n = 3$ ), *H2-Ab1*<sup>-/-</sup> ( $n = 4$ ) and *H2-Ab1*<sup>-/-</sup> *B2m*<sup>-/-</sup> ( $n = 4$ ) mice (c, d); and *Rag1*<sup>-/-</sup> mice ( $n = 6$ ) with adoptive transfer of CD4<sup>+</sup> T cells ( $n = 6$ ; e, f). Results are representative of two independent experiments. Bars indicate mean; \*\*\*\* $P < 0.0001$ , otherwise exact  $P$  values are shown; one-way ANOVA (b, d) or two-tailed Student's  $t$ -test (f). Scale bars, 50  $\mu$ m.

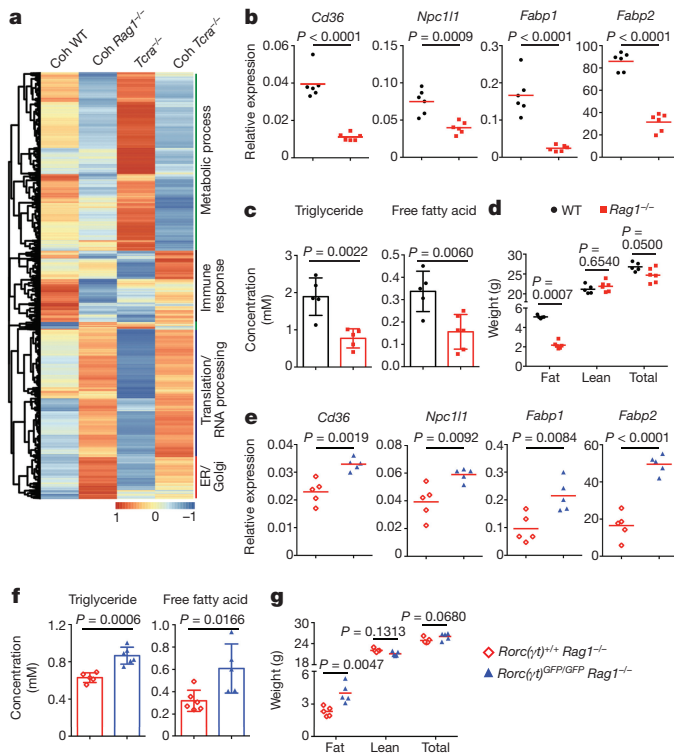
on specific bacteria such as SFB (Fig. 1c), we co-housed wild-type, *Rag1*<sup>-/-</sup>, *Tcr*<sup>-/-</sup> and *Ighm*<sup>-/-</sup> mice for 4–6 weeks and examined STAT3 phosphorylation in these animals. After co-housing, ILC3s and IECs from *Tcr*<sup>-/-</sup> mice but not *Ighm*<sup>-/-</sup> mice showed strong STAT3 phosphorylation (Fig. 3a, b), consistent with a change in the abundance of SFB in the co-housed *Tcr*<sup>-/-</sup> mice (Extended Data Fig. 5c). These data show that  $\alpha\beta$ -T cells are critical for preventing ILC3 activation and IEC signalling induced by SFB. As the presence of SFB was critical for the pSTAT3 signature, all the mice used in the following experiments were co-housed with *Rag1*<sup>-/-</sup> mice unless otherwise noted.

To identify whether CD4<sup>+</sup> helper or CD8<sup>+</sup> cytotoxic cells were responsible for inhibiting ILC3 activation, we examined STAT3 activation in MHCII-deficient (*H2-Ab1*<sup>-/-</sup>), MHCI-deficient (*B2m*<sup>-/-</sup>) and MHCI and MHCII-double-deficient (*H2-Ab1*<sup>-/-</sup> *B2m*<sup>-/-</sup>) mice, which specifically lack CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells or both, respectively. After co-housing with *Rag1*<sup>-/-</sup> mice, both *H2-Ab1*<sup>-/-</sup> and *H2-Ab1*<sup>-/-</sup> *B2m*<sup>-/-</sup> mice had pSTAT3<sup>+</sup> ILC3s and IECs, whereas the equivalent cells in *B2m*<sup>-/-</sup> mice were pSTAT3<sup>-</sup> (Fig. 3c, d). To investigate whether CD4<sup>+</sup> T cells were sufficient to suppress ILC3 activation, we adoptively transferred total CD4<sup>+</sup> T cells into *Rag1*<sup>-/-</sup> mice and



**Figure 4 | Suppression of ILC3 activation by T<sub>reg</sub> and T<sub>H</sub>17 cells.** a–c, Immunofluorescence staining of ileum (a) and quantification of pSTAT3<sup>+</sup> ILC3s (b) and SFB (c) from the small intestine of *Rag1*<sup>-/-</sup> mice ( $n = 7$ ) or *Rag1*<sup>-/-</sup> mice with adoptive transfer of CD4<sup>+</sup> FOXP3-GFP<sup>+</sup> T<sub>reg</sub> cells treated with IL-2-IL-2-antibody complexes ( $n = 7$ ). d, Expression of *Il23p19* in CD64<sup>+</sup> CCR2<sup>-</sup> macrophages ( $n = 5$ ) and CCR2<sup>+</sup> myeloid cells ( $n = 5$ ) from small intestine of mice as in a. e–j, Immunofluorescence staining of ileum (e, h) and quantification of pSTAT3<sup>+</sup> ILC3s (f, i) and SFB (g, j) in distal small intestine of *Rag1*<sup>-/-</sup> mice ( $n = 4$ ) and *Rag1*<sup>-/-</sup> mice with adoptive transfer of 7B8-Tg T cells ( $n = 6$ ) or CBir1-Tg T cells ( $n = 7$ ) (e–g), and *Rag1*<sup>-/-</sup> mice ( $n = 5$ ) and *Rag1*<sup>-/-</sup> mice with adoptive transfer of 7B8-Tg T cells treated with anti-IL-17A antibody ( $n = 5$ ) or control antibody ( $n = 5$ ) (h–j). Results are representative of two independent experiments. Bars show mean (b, c, f, g, i and j) or mean  $\pm$  s.d. (d). \*\*\*\* $P < 0.0001$ , otherwise exact  $P$  values are shown; two-tailed Student's  $t$ -test (b–d) or one-way ANOVA (f, g, i and j). Scale bars, 50  $\mu$ m.

found that they prevented STAT3 phosphorylation in ILC3s and IECs (Fig. 3e, f). Reduction of ILC3 activity by CD4<sup>+</sup> T cells has been reported previously<sup>23</sup>, but the mechanism was not addressed. FOXP3<sup>+</sup> T<sub>reg</sub> and T<sub>H</sub>17 cells are the most abundant subsets of CD4<sup>+</sup> T cells in the intestinal lamina propria<sup>24</sup>. To examine the possible role of T<sub>reg</sub> cells, we isolated these cells from *Foxp3*<sup>GFP</sup> mice and transferred them into *Rag1*<sup>-/-</sup> mice along with IL-2-IL-2-antibody complexes to maintain their number and function<sup>10,25</sup>. Six to eight weeks after transfer, SFB abundance in the small intestine of these mice had not changed, but pSTAT3 in ILC3s and IECs was diminished (Fig. 4a–c). This effect was concordant with reduced *Il23a* expression in CCR2<sup>+</sup> myeloid-cell populations (Fig. 4d). SFB preferentially induces antigen-specific effector CD4<sup>+</sup> T<sub>H</sub>17 cells<sup>19,20</sup>. To examine the role of T<sub>H</sub>17 cells in controlling ILC3 activation, we took advantage of SFB-specific T cell receptor transgenic mice (7B8-Tg). Adoptive transfer of naive 7B8-Tg CD4<sup>+</sup> T cells into *Rag1*<sup>-/-</sup> recipients markedly decreased SFB abundance in the small intestine, limiting ILC3 activation. By comparison, CBir1-Tg CD4<sup>+</sup> T cells, which recognize commensal-derived flagellin, had only a modest effect on the abundance of SFB and ILC3 activation (Fig. 4e–g). Treatment with IL-17A-neutralizing antibodies partially reversed the effect of 7B8-Tg cells on ILC3 activation and SFB reduction; the incomplete nature of this effect is likely to be due to activities of other cytokines such as IL-17F (Fig. 4h, i). These data indicate that both T<sub>reg</sub> and T<sub>H</sub>17 cells contribute to preventing ILC3 activation: T<sub>reg</sub> cells do so by decreasing IL-23 production from CCR2<sup>+</sup> monocytes and monocyte-derived dendritic cells, whereas effector T cells regulate the bacterial burden. Both of these mechanisms prevent the activation of the IL-23-ILC3-IL-22-IEC circuit.



**Figure 5 | Disrupted lipid metabolism in ILC3 and IEC-activated mice.** **a**, RNA-seq analysis of ileal tissue from co-housed (Coh) wild-type, *Rag1*<sup>-/-</sup> and *Tcra*<sup>-/-</sup> mice and non-co-housed *Tcra*<sup>-/-</sup> mice (*n* = 3). **b–g**, Expression of indicated genes (**b, e**), serum triglyceride and free fatty acid levels (**c, f**), and body composition (**d, g**) of wild-type (*n* = 5) or *Rag1*<sup>-/-</sup> mice (*n* = 6) (**b–d**), and *Rorc*( $\gamma$ )<sup>+/+</sup> *Rag1*<sup>-/-</sup> (*n* = 5) or *Rorc*( $\gamma$ )<sup>GFP/GFP</sup> *Rag1*<sup>-/-</sup> mice (*n* = 5) (**e–g**). Results are representative of two independent experiments. Bars show mean (**b, d, e** and **g**) or mean  $\pm$  s.d. (**c** and **f**). Exact *P* values are given and calculated by two-tailed Student's *t*-test.

The evidence that ILC3s and adaptive lymphocytes operate sequentially and use markedly distinct effector mechanisms in their interactions with commensal bacteria raised the question of whether persistent ILC3 activity in the absence of adaptive immune control might have a negative effect on host homeostasis. To examine this issue, we performed whole-tissue RNA sequencing (RNA-seq) analysis of ilea from co-housed wild-type, *Rag1*<sup>-/-</sup> and *Tcra*<sup>-/-</sup> mice and separately housed *Tcra*<sup>-/-</sup> mice. Genes that encode cytokines and anti-microbial peptides involved in microbial control were expressed at higher levels in pSTAT3<sup>+</sup> *Rag1*<sup>-/-</sup> and co-housed *Tcra*<sup>-/-</sup> mice. By contrast, metabolic processing genes showed substantially lower expression in pSTAT3<sup>+</sup> mice (Fig. 5a). Using quantitative real-time PCR (qPCR) analysis, we confirmed the reduction in mRNA coding for key lipid transporters, including *Cd36*, *Npc1l1*, *Fabp1* and *Fabp2*, in *Rag1*<sup>-/-</sup> and co-housed *Tcra*<sup>-/-</sup> mice (Fig. 5b and Extended Data Fig. 6a). This change was associated with significantly decreased serum levels of triglycerides and free fatty acids (Fig. 5c and Extended Data Fig. 6b). At the macroscopic level, *Rag1*<sup>-/-</sup> mice exhibited less fat accumulation than wild-type mice that were fed a standard chow diet (Fig. 5d). Elimination of ILC3s in *Rorc*( $\gamma$ )<sup>GFP/GFP</sup> *Rag1*<sup>-/-</sup> mice had the opposite effect, with higher levels of lipid transporters, serum triglycerides and free fatty acids, and higher fat storage than in *Rag1*<sup>-/-</sup> mice with activated ILC3s (Fig. 5e–g). To investigate whether these lipid abnormalities were specifically associated with persistent IL-22 production, we administered adenoviruses expressing IL-22 or GFP to wild-type and non-co-housed *Tcra*<sup>-/-</sup> mice. Injection of adenovirus expressing IL-22 induced notable STAT3 activation in IECs of wild-type mice, comparable to that seen in *Rag1*<sup>-/-</sup> mice (Extended Data Fig. 7), and led to reduced expression of lipid transporters in the gut and lower

serum lipid concentration (Extended Data Fig. 6c–f). These findings suggest that, in the absence of adaptive lymphocytes, although ILCs are capable of constraining microbial communities their persistent activation results in abnormal lipid handling and tissue homeostasis. Previous studies suggested a role for IL-22 in lipid metabolism<sup>26</sup>, with a recent report indicating that IL-22 promoted lipid transporter expression in IECs in the small intestine<sup>27</sup>. It is likely that transient, low-level IL-22 production<sup>27</sup> has a notably different effect than the high-level, persistent exposure of IECs to this cytokine that we report here.

Previous studies of ILCs have focused on the roles of these cells in resistance to pathogens or in immunopathological conditions such as chronic inflammatory diseases<sup>6</sup>, with recent studies suggesting that they also have a role in neurobiological function<sup>28–30</sup>. Here we demonstrate a clear role for these cells in the establishment of a compatible commensal state. Early in life, they temper the expansion of some bacterial species, especially those with inflammatory potential, protecting the epithelium of the gut. As the adaptive immune system matures, CD4<sup>+</sup> T cells respond to the commensal population, and through mechanisms other than the IL-23-induced IL-22 pathway used by the ILC3s, establish a state of non-inflammatory commensalism in which the ILCs are largely quiescent. In the absence of a dominant adaptive immune response, the persistent activation of ILC3s results in impaired lipid metabolism. Our findings may have bearing on recent studies that show that a failure to develop a ‘mature’ gut microbiota is associated with severe nutritional abnormalities and other pathological states in humans as well as in germ-free animals colonized with microbial material from individuals with this condition<sup>31,32</sup>. Taken together, our results increase understanding of how the innate and adaptive immune systems act sequentially on the developing gut microbiota to establish a balanced commensal state that supports normal tissue function.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 September 2016; accepted 6 December 2017.

Published online 22 January 2018.

- Caballero, S. & Pamer, E. G. Microbiota-mediated inflammation and antimicrobial defense in the intestine. *Annu. Rev. Immunol.* **33**, 227–256 (2015).
- Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
- Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
- Renz, H., Brandtzaeg, P. & Hornef, M. The impact of perinatal immune development on mucosal homeostasis and chronic inflammation. *Nat. Rev. Immunol.* **12**, 9–23 (2011).
- Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science* **352**, 539–544 (2016).
- Artis, D. & Spits, H. The biology of innate lymphoid cells. *Nature* **517**, 293–301 (2015).
- Bando, J. K. & Colonna, M. Innate lymphoid cell function in the context of adaptive immunity. *Nat. Immunol.* **17**, 783–789 (2016).
- Gerner, M. Y., Kastenmuller, W., Ifrim, I., Kabat, J. & Germain, R. N. Histocytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Immunity* **37**, 364–376 (2012).
- Gerner, M. Y., Torabi-Parizi, P. & Germain, R. N. Strategically localized dendritic cells promote rapid T cell responses to lymph-borne particulate antigens. *Immunity* **42**, 172–185 (2015).
- Liu, Z. *et al.* Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature* **528**, 225–230 (2015).
- Nguyen, P. M., Putoczki, T. L. & Ernst, M. STAT3-activating cytokines: a therapeutic opportunity for inflammatory bowel disease? *J. Interferon Cytokine Res.* **35**, 340–350 (2015).
- Klose, C. S. *et al.* A T-bet gradient controls the fate and function of CCR6<sup>+</sup> ROR $\gamma$ <sup>+</sup> innate lymphoid cells. *Nature* **494**, 261–265 (2013).
- Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
- Sano, T. *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector Th17 responses. *Cell* **163**, 381–393 (2015).
- Guo, X. *et al.* Induction of innate lymphoid cell-derived interleukin-22 by the transcription factor STAT3 mediates protection against intestinal infection. *Immunity* **40**, 25–39 (2014).

16. Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
17. Tamoutounour, S. *et al.* Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39**, 925–938 (2013).
18. Jiang, H. Q., Bos, N. A. & Cebra, J. J. Timing, localization, and persistence of colonization by segmented filamentous bacteria in the neonatal mouse gut depend on immune status of mothers and pups. *Infect. Immun.* **69**, 3611–3617 (2001).
19. Goto, Y. *et al.* Segmented filamentous bacteria antigens presented by intestinal dendritic cells drive mucosal T<sub>H</sub>17 cell differentiation. *Immunity* **40**, 594–607 (2014).
20. Yang, Y. *et al.* Focused specificity of intestinal T<sub>H</sub>17 cells towards commensal bacterial antigens. *Nature* **510**, 152–156 (2014).
21. Eberl, G., Di Santo, J. P. & Vivier, E. The brave new world of innate lymphoid cells. *Nat. Immunol.* **16**, 1–5 (2015).
22. Fang, D. & Zhu, J. Dynamic balance between master transcription factors determines the fates and functions of CD4 T cell and innate lymphoid cell subsets. *J. Exp. Med.* **214**, 1861–1876 (2017).
23. Korn, L. L. *et al.* Conventional CD4<sup>+</sup> T cells regulate IL-22-producing intestinal innate lymphoid cells. *Mucosal Immunol.* **7**, 1045–1057 (2014).
24. Littman, D. R. & Rudensky, A. Y. T<sub>H</sub>17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140**, 845–858 (2010).
25. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in FOXP3-expressing regulatory T cells. *Nat. Immunol.* **6**, 1142–1151 (2005).
26. Wang, X. *et al.* Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes. *Nature* **514**, 237–241 (2014). 10.1038/nature13564
27. Wang, Y. *et al.* The intestinal microbiota regulates body composition through NFIL3 and the circadian clock. *Science* **357**, 912–916 (2017).
28. Ibiza, S. *et al.* Glial-cell-derived neuroregulators control type 3 innate lymphoid cells and gut defence. *Nature* **535**, 440–443 (2016).
29. Cardoso, V. *et al.* Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. *Nature* **549**, 277–281 (2017).
30. Klose, C. S. N. *et al.* The neuropeptide neuromedin U stimulates innate lymphoid cells and type 2 inflammation. *Nature* **549**, 282–286 (2017).
31. Blanton, L. V., Barratt, M. J., Charbonneau, M. R., Ahmed, T. & Gordon, J. I. Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics. *Science* **352**, 1533 (2016).
32. Subramanian, S. *et al.* Cultivating healthy growth and nutrition through the gut microbiota. *Cell* **161**, 36–48 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank Y. Choi and D.M. Kobuley for providing germ free *Rag1*<sup>−/−</sup> mice and performing SFB mono-colonization; M. Oukka and S. K. Durum for providing mice; M. Mack, B. Gao and Y. Umesaki for providing anti-CCR2 antibody, IL-22 adenovirus and SFB faecal pellets; C. Eigsti, V. Nair and J. Davis for cell sorting, scanning electron microscopy and microbiota analysis; J. Zhu for discussions; and members of the Laboratory of Systems Biology for their comments during the course of these studies and input during preparation of this manuscript. Y.H. was supported by an NIAID K99 award (1K99AI123350-01A1). This research was supported by the Intramural Research Program of NIAID, NIH.

**Author Contributions** K.M. designed and conducted most of the experiments and data analysis and prepared the manuscript; A.P.B., S.T. and Y.H. helped with cell isolation and transfer; L.Z. measured mouse body composition; N.B. performed the analysis of microbiota translocation; A.J.M. performed the RNA-seq and data analysis; M.Y.G. provided helpful suggestions regarding imaging and histo-cytometry; Y.B. provided helpful suggestions, discussed data interpretation and contributed to the manuscript; and R.N.G. designed experiments, interpreted data and helped to write the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to K.M. ([kairui.mao@nih.gov](mailto:kairui.mao@nih.gov)) and R.N.G. ([rgermain@nih.gov](mailto:rgermain@nih.gov)).

## METHODS

**Data reporting.** Mice of similar ages were randomly allocated into different groups. For most experiments (co-housing, antibody treatment and cell transfer experiments), mice were ear-tagged with numbers and investigators did not know the identity of the specific samples until after data were analysed.

**Mice.** C57BL/6, *Rag1*<sup>−/−</sup>, *Il23a*<sup>−/−</sup>, *Foxp3*<sup>GFP</sup>, *Tcrα*<sup>−/−</sup>, *Ighm*<sup>−/−</sup>, *B2m*<sup>−/−</sup>, *H2-Ab1*<sup>−/−</sup> and *H2-Ab1*<sup>−/−</sup>*B2m*<sup>−/−</sup> mice were obtained from Taconic Laboratories through a special NIAID contract. *Rorc*( $\gamma$ t)<sup>GFP/GFP</sup>, *Il6*<sup>−/−</sup> and 7B8-Tg mice were purchased from Jackson Laboratories. *Il22*-tdTomato mice were provided by S. K. Durum. *Il22*<sup>−/−</sup> mice were provided by Genentech. Unless specified, mice used in the study were eight- to sixteen-week-old males. All mice were maintained in specific-pathogen-free conditions at an Association for Assessment and Accreditation of Laboratory Animal Care-accredited animal facility at the NIAID and were used under a study protocol approved by NIAID Animal Care and Use Committee (National Institutes of Health).

Germ-free *Rag1*<sup>−/−</sup> mice were provided by the Penn Gnotobiotic Mouse Facility and experiments involving these mice were performed at the Penn Gnotobiotic Mouse Facility. Germ-free mice were maintained in sterile plastic isolator units and fed autoclaved LabDiet5021 mouse chow (LabDiet) and autoclaved water.

**Immunofluorescence staining and confocal imaging.** The ileal portion of the small intestine was excised and prepared using the Swiss roll technique, then incubated in a fixation and permeabilization solution (BD Bioscience, 554722) overnight followed by dehydration in 30% sucrose before embedding in OCT compound (Sakura Finetek). 18- $\mu$ m sections were cut on a CM3050S cryostat (Leica) and adhered to Superfrost Plus slides (VWR). Frozen sections were treated with methanol for 20 min at −20°C and then permeabilized and blocked in PBS containing 0.3% Triton X-100 (Sigma-Aldrich) and 10% normal mouse serum (Jackson Immunoresearch) followed by staining with antibodies diluted in blocking buffer. The following antibodies were used for staining: anti-CD3 (17A2; Biolegend), anti-EpCAM (G8.8, Biolegend), anti-ROR $\gamma$ t (AFKJS-9, eBioscience), anti-CD90.2 (30-H12, Biolegend) and anti-pSTAT3 (D3A7, Cell Signaling Technology). After staining, slides were mounted with Fluormount G (Southern Biotech), and examined on a Leica TCS SP8 confocal microscope. Images were analysed with Imaris software (Bitplane).

**Antibiotic treatment.** Male six-week-old *Rag1*<sup>−/−</sup> mice were provided with ampicillin (1 g/l), kanamycin (5 g/l), vancomycin (500 mg/l), neomycin trisulfate (1 g/l) and metronidazole (1 g/l) in drinking water for three weeks. All antibiotics were purchased from Sigma-Aldrich.

**Histo-cytometry.** Histo-cytometry analysis was performed as previously described<sup>8–10</sup>, with minor modifications. In brief, multi-parameter confocal images were corrected for fluorophore spillover using the Leica Channel Dye Separation module. For analysis of pSTAT3<sup>+</sup> ILC3s, the ILC3 surface was constructed on the basis of the ROR $\gamma$ t channel, and the object statistics were exported into FlowJo X (TreeStar) for analysis and graphing with Prism (GraphPad).

**Co-housing and mono-colonization with SFB.** Wild-type age-matched male mice were ear-tagged and housed in a cage with an equal number of the respective knockout mice for 4–8 weeks. For association of germ-free *Rag1*<sup>−/−</sup> mice with SFB (a gift from Y. Umesaki), faecal pellets isolated from SFB mono-associated mice were reconstituted in sterile PBS and 200  $\mu$ l of this suspension was administered to each germ-free mouse by gavage in a sterile isolator. SFB reconstitution was confirmed by qPCR of faecal 16S ribosomal DNA relative to negative germ-free controls as previously described<sup>33</sup>. Mono-associated mice were maintained for 3–4 weeks before analysis.

**Lamina propria myeloid cell isolation.** Small intestinal segments were treated with medium containing 5 mM EDTA and 0.145 mg/ml dithiothreitol for 30 min at 37°C with constant stirring. Tissue was further digested with 100  $\mu$ g/ml Liberase TL (Roche) and 500  $\mu$ g/ml DNase I (Sigma-Aldrich), with continuous stirring at 37°C for 30 min. Digested tissue was forced through a Collector tissue sieve (Bellco Glass) and passed through 70- and 40- $\mu$ m cell strainers. Cells were washed and incubated with a mixture of monoclonal antibodies containing anti-CD11c (N418, Biolegend), anti-MHC II (M5/114.15.2, eBioscience), anti-CD45 (30-F11, BD Bioscience), anti-CD24 (M1/69, Biolegend), anti-CD64 (X54-5/7.1, Biolegend), anti-CCR2 (#475301, R&D Systems), anti-CD11b (M1/70, eBioscience), anti-Ly-6C (HK1.4, Biolegend) and anti-CD103 (2E7, eBioscience), as well as monoclonal antibodies against the non-dendritic-cell components: anti-Ly-6G (1A8, BD Biosciences), anti-NK1.1 (PK136, BD Biosciences), anti-TCR  $\beta$  (H57-597, BD Biosciences), anti-TCR  $\gamma\delta$  (GL3, BD Biosciences), anti-Siglec F (E50-2440, BD Biosciences) and anti-B220 (RA3-6B2, BD Biosciences). Different myeloid-cell populations were sorted by flow cytometry on a FACSria. Purity was verified by flow cytometry on a FACSria. The purity of all populations was > 99%.

**Adoptive transfer and antibody treatment.** CD4<sup>+</sup> T cells were isolated from lymph nodes and spleens from wild-type or *Foxp3*<sup>GFP</sup> mice using the CD4<sup>+</sup>

T Cell Isolation Kit (Miltenyi, 130-104-454). T<sub>reg</sub> cells were sorted according to GFP expression, and  $1 \times 10^7$  total CD4<sup>+</sup> T cells or  $1 \times 10^6$  T<sub>reg</sub> cells were transferred into *Rag1*<sup>−/−</sup> mice intravenously. Naive T cells from 7B8-Tg mice or from CBir1-Tg mice were isolated using the naive CD4<sup>+</sup> T cell Kit (Miltenyi, 130-104-453) and transferred into *Rag1*<sup>−/−</sup> mice intravenously ( $1 \times 10^6$  cells per mouse) for 6–8 weeks.

IL-2–IL-2-antibody complexes were made by mixing 1  $\mu$ g recombinant mouse IL-2 (Biolegend) with 5  $\mu$ g anti-IL-2 monoclonal antibody (clone JES6-1, BioXcell) followed by incubation at 37°C for 30 min. Mice were injected intraperitoneally with IL-2–IL-2-antibody complexes, 150  $\mu$ g anti-IL-17A (Clone 17F3, BioXcell) or isotype control IgG twice a week after T-cell transfer. For monocyte depletion, *Rag1*<sup>−/−</sup> mice were injected intraperitoneally with 20  $\mu$ g anti-CCR2 (clone MC-21, a gift from M. Mack), 200  $\mu$ g anti-Gr1 (clone RB6-8C5, BioXcell) or isotype control IgG on days 0, 1, 2, 3, 5, 7, 9, 11 and 13. Mice were euthanized 24 h after the last injection.

**Quantitative real-time PCR.** RNA from the terminal ileum was isolated with Trizol reagent (Thermo Fisher Scientific). cDNA was synthesized using the iScript cDNA Synthesis kit (Bio-Rad) according to the manufacturer's instructions. qPCR was performed using SsoFast EvaGreen Supermixes (Bio-Rad) or LightCycler 480 Probes Master (Roche). Reactions were run with the CFX Connect Real-Time PCR Detection System (Bio-Rad). The primer and probe set for *mIl23p19* (Mm00518984) was purchased from Thermo Fisher Scientific. See Extended Data Table 1 for a list of primers and probes used in this study.

**Analysis of microbiota in the small intestine.** The contents of the small intestine were collected from different mice, and bacterial DNA was extracted with the QIAamp DNA Stool Kit (QIAGEN) according to the manufacturer's instructions. Different species of bacteria were quantified by qPCR with primers specific for 16S rRNA genes using SsoFast EvaGreen Supermixes (Bio-Rad). See Extended Data Table 1 for a list of primers.

**Scanning electron microscopy.** Mouse terminal ilea (immediately proximal to the ileal–caecal junction) were collected, slit and fixed overnight in fresh fixative. All subsequent processing was carried out in a Pelco Biowave microwave (TedPella) at 250 W under 15-inch Hg vacuum. The tissue pieces were washed in 0.1 M sodium cacodylate buffer and post-fixed in 1% osmium tetroxide reduced with 0.8% potassium ferrocyanide in 0.1 M sodium cacodylate buffer. After three washes with buffer, the tissue pieces were dehydrated in a graded ethanol series and dried using a Bal-Tec 030 critical point dryer. The tissue pieces were subsequently coated with 70 Å iridium using ion-beam sputtering (South Bay Technology) and imaged using a SU-8000 scanning electron microscope (Hitachi).

**RNA-seq analysis.** Total RNA was extracted from distal small intestine using TRIzol (Thermo Fisher Scientific). RNA-seq libraries were prepared using the Illumina TruSeq Stranded LT mRNA library preparation kit (Illumina), and libraries were sequenced on an Illumina NextSeq 500 to a read depth of 25–35 million reads per sample. Sequencing quality control was assessed using FastQC version 0.11.3 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped to the UCSC mm9 *Mus musculus* genome using Tophat2<sup>34</sup> with the following flags: –library-type fr-firststrand –r 25. The tool featureCounts from the subread<sup>35</sup> Python package was used to count reads mapping to gene features, and the R package DESeq2\_ENREF\_38<sup>36</sup> was used to test for differential expression and to generate regularized log-transformed (rlog) count tables. The R package pheatmap was used to generate the heatmap of the RNA-seq results, and ggplot2 was used to generate scatterplots of the rlog counts.

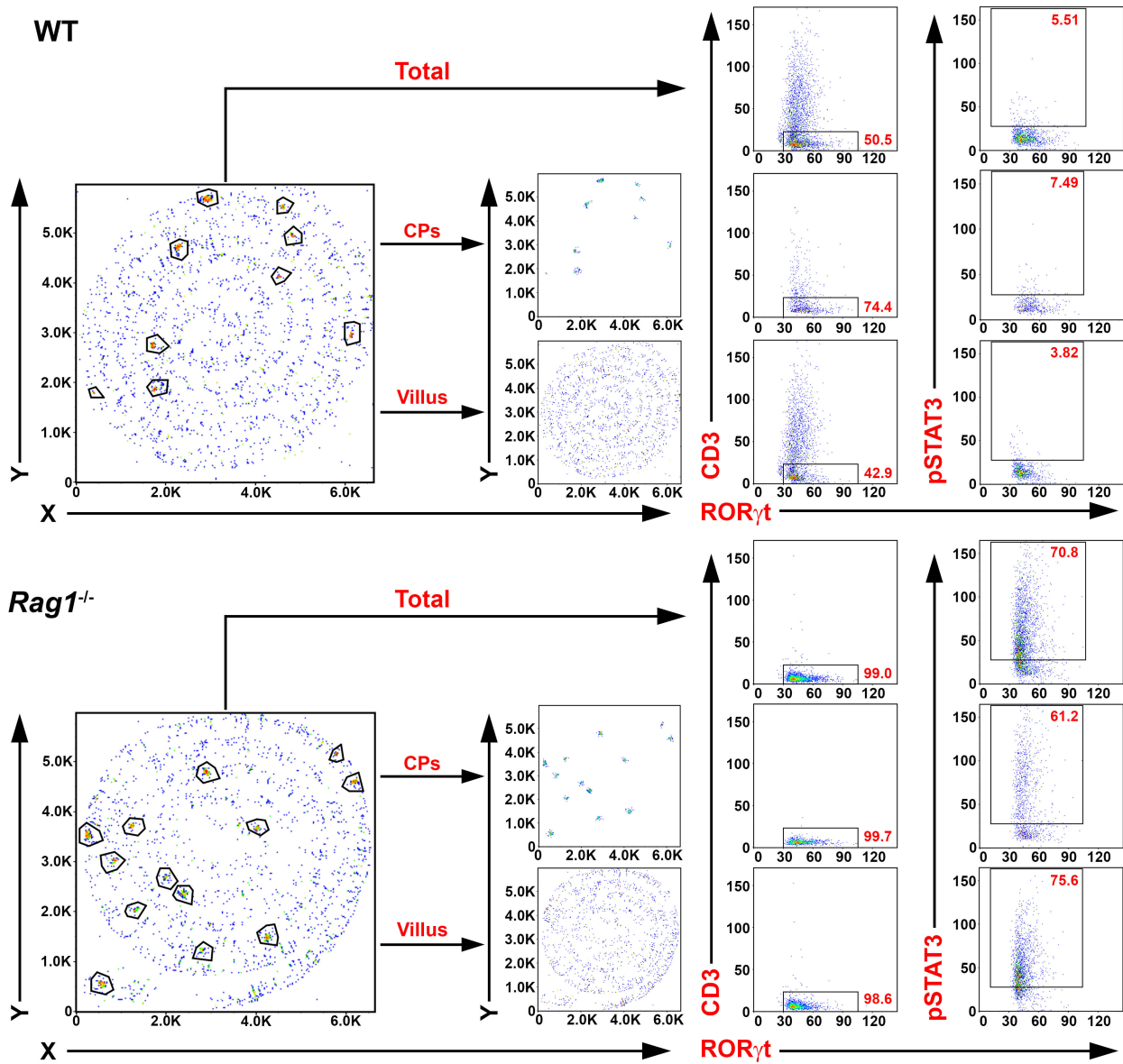
**Serum triglycerides, free fatty acids and body composition measurement.** Serum triglycerides and free fatty acids were measured using a Triglyceride Reagent Set (Pointe Scientific, Inc.) and Free Fatty Acid Quantitation Kit (Sigma-Aldrich) according to the manufacturer's instructions. Mouse body composition, including fat and lean masses, was measured with EchoMRI at four months of age.

**Administration of mice with IL-22 adenovirus.** Adenoviruses expressing IL-22 and GFP were provided by B. Gao. IL-22 adenovirus was made by cloning mouse IL-22 cDNA (544 bp) into the pENTR/D-TOPO system (Invitrogen), followed by using the Gateway system (Invitrogen) to perform an LR reaction with pAd/CMV/V5-DEST to make the expression vector pAd/CMV/mIL-22. Mice were injected intravenously with  $2 \times 10^8$  pfu IL-22 adenovirus or GFP adenovirus.

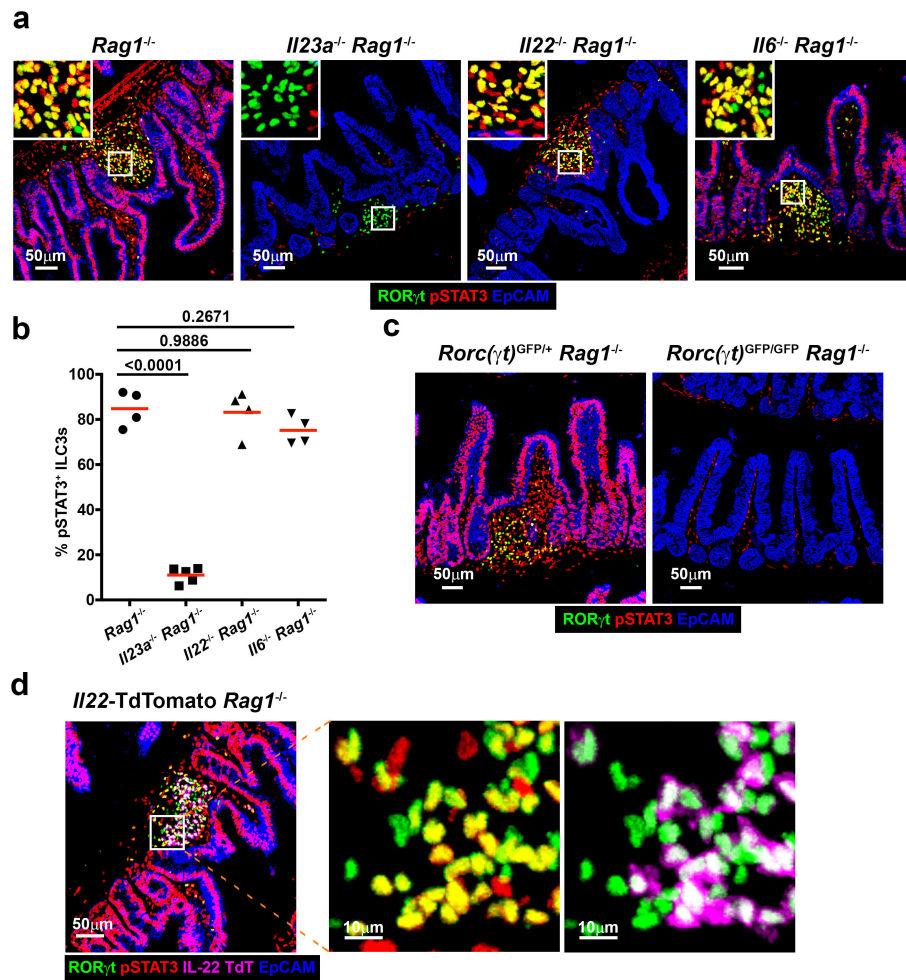
**Statistical analysis.** No statistical methods were used to predetermine sample size. Prism software (GraphPad) was used for all statistical analysis. Student's *t*-test (two-tailed), or one-way or two-way ANOVA were used for the statistical analysis of differences between two groups; \*\*\**P* < 0.0001, and exact *P* values are shown in figures.

**Data availability.** RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus database with the accession number GSE86780. Figure 5a shows RNA-seq-related data. The data that support the findings of this study are available from the corresponding author upon reasonable request.

33. Qiu, J. *et al.* Group 3 innate lymphoid cells inhibit T-cell-mediated intestinal inflammation through aryl hydrocarbon receptor signaling and regulation of microflora. *Immunity* **39**, 386–399 (2013).
34. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
35. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

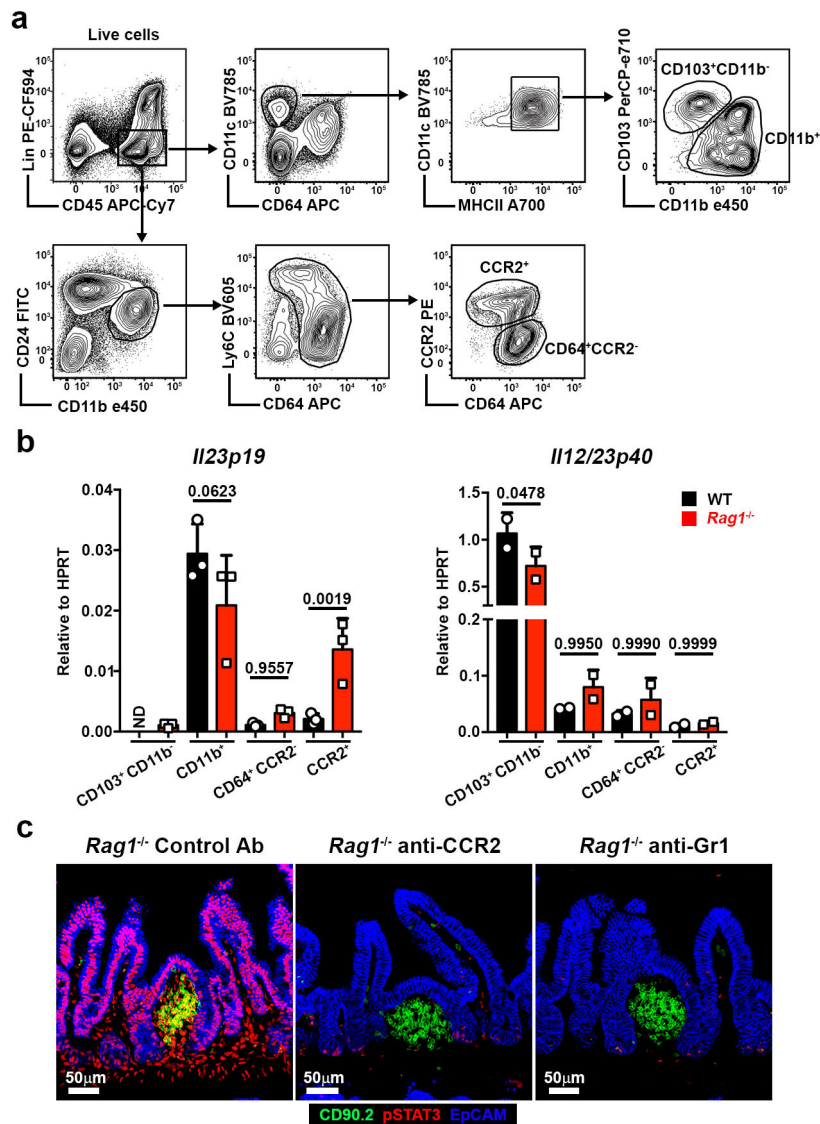


**Extended Data Figure 1 | Quantification of pSTAT3<sup>+</sup> ILC3s by histo-cytometry.** Gating strategy for analysis of pSTAT3<sup>+</sup> ILC3s from small intestine of wild-type or Rag1<sup>-/-</sup> mice.



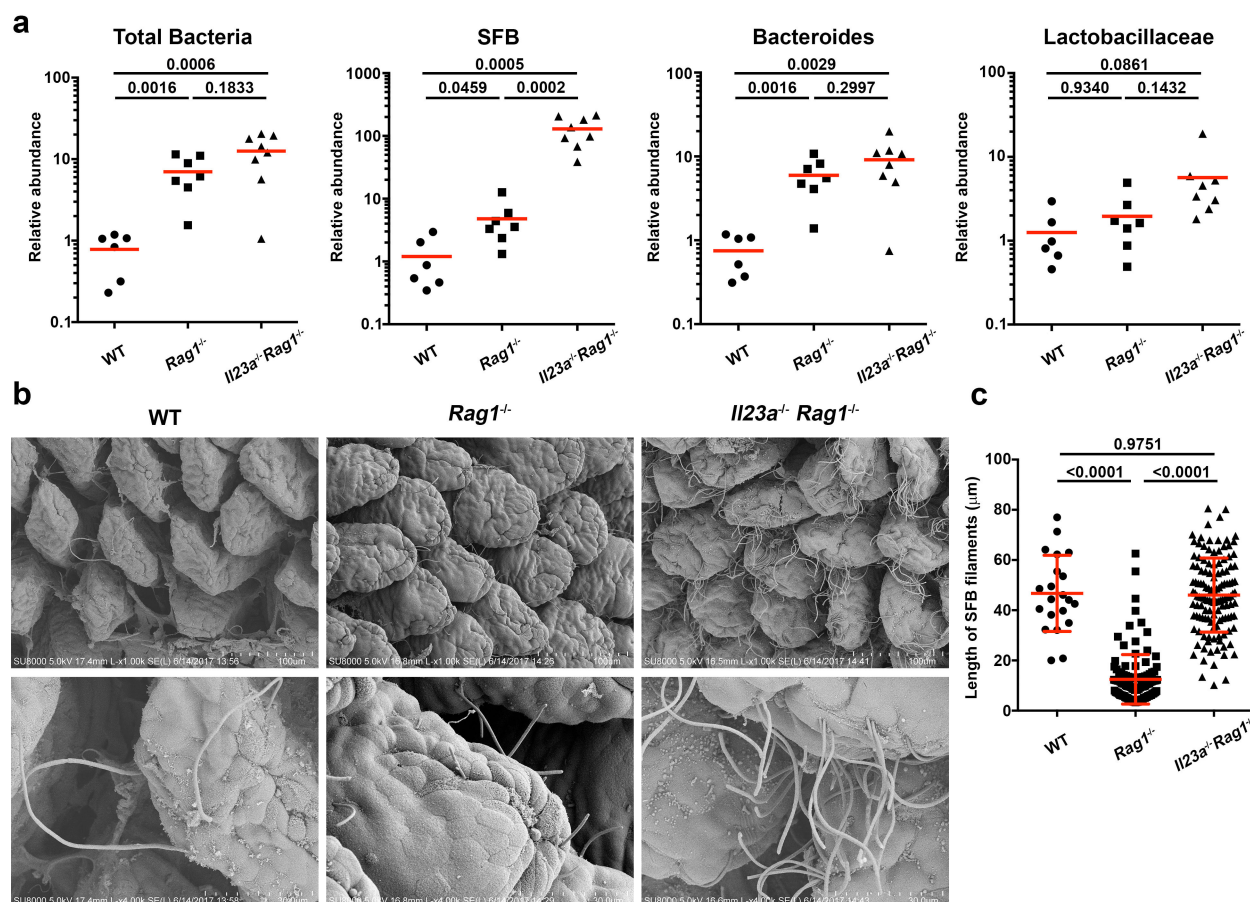
**Extended Data Figure 2 | Cellular and molecular mechanism of STAT3 activation in *Rag1*<sup>-/-</sup> small intestine.** **a**, Immunofluorescence staining of ileum from *Rag1*<sup>-/-</sup> ( $n=4$ ), *Il23a*<sup>-/-</sup>*Rag1*<sup>-/-</sup> ( $n=5$ ), *Il22*<sup>-/-</sup>*Rag1*<sup>-/-</sup> ( $n=4$ ) and *Il6*<sup>-/-</sup>*Rag1*<sup>-/-</sup> mice ( $n=4$ ). **b**, Percentage of pSTAT3<sup>+</sup> ILC3s in **a**. **c–d**, Immunofluorescence staining of ileum from

*Rorc*( $\gamma$ t)<sup>GFP/+</sup>*Rag1*<sup>-/-</sup> and *Rorc*( $\gamma$ t)<sup>GFP/GFP</sup>*Rag1*<sup>-/-</sup> mice ( $n=4$ ; **c**) and *Il22*-tdTomato *Rag1*<sup>-/-</sup> mice ( $n=3$ ; **d**). Results are representative of three independent experiments. Bars show mean; exact  $P$  values are given and calculated by one-way ANOVA.



**Extended Data Figure 3 | Mononuclear phagocyte subpopulation responsible for IL-23 production and pSTAT3 activation.** **a**, Flow cytometry of total live cells from the small intestine lamina propria, showing the gating strategy for sorting different myeloid-cell subsets: CD103<sup>+</sup>CD11b<sup>-</sup> and CD11b<sup>+</sup> conventional dendritic cells, CD64<sup>+</sup>CCR2<sup>-</sup> macrophages and CCR2<sup>+</sup> monocytes and monocyte-derived dendritic cells.

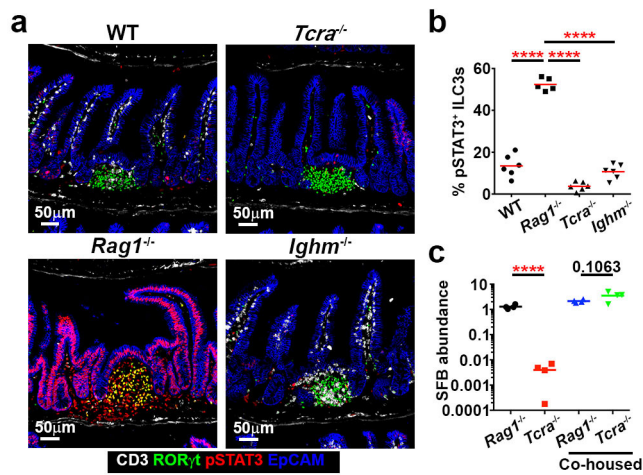
**b**, Expression of *Il23a* ( $n = 3$ ) and *Il12b* (which encodes the p40 subunit of IL-12) in different cell populations sorted as in **a** ( $n = 2$ ). **c**, Immunofluorescence staining of ileum from *Rag1*<sup>-/-</sup> mice and *Rag1*<sup>-/-</sup> mice treated with anti-CCR2 or anti-Gr1 antibody for two weeks ( $n = 4$ ). Results are representative of three independent experiments. Mean  $\pm$  s.d.; exact *P* values are given and calculated by two-way ANOVA.



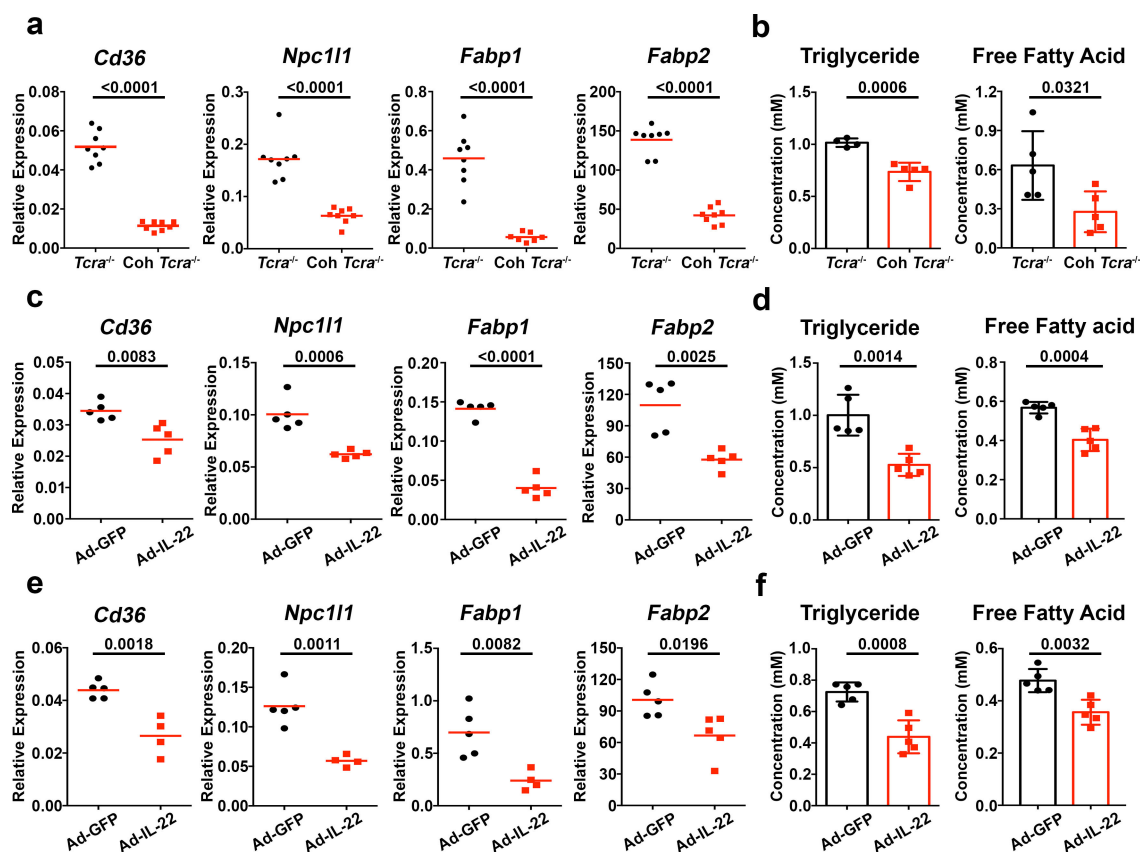
**Extended Data Figure 4 | Microorganisms in the small intestine of co-housed wild-type,  $Rag1^{-/-}$  and  $Il23a^{-/-} Rag1^{-/-}$  mice.**

**a**, Quantification of indicated bacteria species in the ileum of co-housed wild-type ( $n=6$ ),  $Rag1^{-/-}$  ( $n=7$ ) and  $Il23a^{-/-} Rag1^{-/-}$  ( $n=8$ ) mice by real-time PCR with primers specific to 16S rRNA genes. Results are pooled

from two independent experiments. **b**, Scanning electron microscopy of terminal ileum of co-housed mice as in **a** ( $n=3$ ). **c**, Quantification of the length of SFB filaments in **b**. Results are representative of two independent experiments. Bars show mean (**a**) and mean  $\pm$  s.d. (**c**); exact  $P$  values are given and calculated by one-way ANOVA.

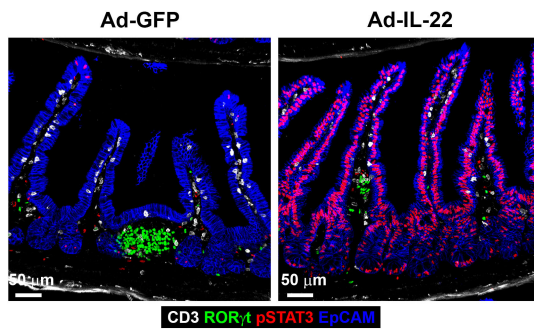


**Extended Data Figure 5 | Lack of ILC3 activation in SFB negative *Tcra*<sup>-/-</sup> mice.** **a**, Immunofluorescence staining of ileum from wild-type ( $n=6$ ), *Rag1*<sup>-/-</sup> ( $n=5$ ), *Tcra*<sup>-/-</sup> ( $n=5$ ) and *Ighm*<sup>-/-</sup> mice ( $n=6$ ). **b**, Percentage of pSTAT3<sup>+</sup> ILC3s in **a**. **c**, Quantification of SFB in distal small intestine of non-co-housed *Rag1*<sup>-/-</sup> and *Tcra*<sup>-/-</sup> or co-housed *Rag1*<sup>-/-</sup> and *Tcra*<sup>-/-</sup> mice ( $n=4$ ). Results are representative of three (**a**, **b**) or two (**c**) independent experiments. Bars show mean; \*\*\*\* $P < 0.0001$ ; otherwise exact  $P$  values are shown and calculated by one-way ANOVA.



**Extended Data Figure 6 | Dysregulation of lipid metabolism by IL-22.** **a**, Expression of indicated genes in the ileum from age-matched male *Tcra*<sup>-/-</sup> mice and *Tcra*<sup>-/-</sup> mice co-housed with *Rag1*<sup>-/-</sup> mice (*n* = 8). Results are pooled from two independent experiments. **b**, Serum triglyceride and free fatty acid levels from *Tcra*<sup>-/-</sup> mice and *Tcra*<sup>-/-</sup> mice co-housed with *Rag1*<sup>-/-</sup> mice (*n* = 5). **c–f**, Expression of indicated genes in

the ileum (c, e) and serum triglyceride and free fatty acid levels (d, f) from wild-type (c, d) or *Tcra*<sup>-/-</sup> mice (e, f) injected with adenovirus expressing IL-22 (*n* = 5) or GFP (*n* = 5). Results are representative of two independent experiments. Bars show mean (a, c, e) and mean ± s.d. (b, d, f); exact *P* values are given and calculated by two-tailed Student's *t*-test.



**Extended Data Figure 7 | STAT3 activation in IECs by IL-22 adenovirus.** Immunofluorescence staining of ileum from C57BL/6 mice injected with adenoviruses expressing either IL-22 or GFP for two weeks. Images are representative of three sections from three mice of each group and results are representative of two independent experiments.

Extended Data Table 1 | Primers and probes for quantitative PCR

|                       |                              |
|-----------------------|------------------------------|
| <i>Hprt</i> FW        | TGAAGAGCTACTGTAATGATCAGTCA   |
| <i>Hprt</i> RV        | AGCAAGCTTGCAACCTTAACCA       |
| <i>Il22</i> FW        | CATGCAGGAGGTGGTACCTT         |
| <i>Il22</i> RV        | CAGACGCAAGCATTCTCAG          |
| <i>Reg3b</i> FW       | ATGGCTCCTACTGCTATGCC         |
| <i>Reg3b</i> RV       | GTGTCCTCCAGGCCTCTT           |
| <i>Reg3g</i> FW       | CAAGGTGAAGTTGCCAAGAA         |
| <i>Reg3g</i> RV       | CCTCTGTTGGGTTCATAGCC         |
| <i>SAA1/2</i> FW      | CTGCCTGCCAAATACTGAGAGTC      |
| <i>SAA1/2</i> RV      | CCACTTCCAAGTTCCTGTTTATTAC    |
| <i>Eubacteria</i> FW  | ACTCCTACGGGAGGCAGCAGT        |
| <i>Eubacteria</i> RV  | ATTACCGCGGCTGCTGGC           |
| <i>SFB</i> FW         | GACGCTGAGGCATGAGAGCAT        |
| <i>SFB</i> RV         | GACGGCACGGATTGTTATTCA        |
| <i>Bacteroides</i> FW | GGTTCGAGAGGAGGTCCC           |
| <i>Bacteroides</i> RV | GCTGCCTCCCGTAGGAGT           |
| <i>Hprt</i> FW        | TGGATATGCCCTTGACTATAATGAG    |
| <i>Hprt</i> RV        | TGGCAACATCAACAGGACTC         |
| <i>Hprt</i> Probe     | TCAACTTGCGCTCATCTTAGGCTTTGTA |
| <i>Cd36</i> FW        | GCGACATGATTAATGGCACAG        |
| <i>CD36</i> RV        | GATCCGAACACAGCGTAGATAG       |
| <i>CD36</i> Probe     | CAACAAAAGGTGGAAAGGAGGCTGC    |
| <i>Npc1l1</i> FW      | CGGAACTCACAGGACTTTACAG       |
| <i>Npc1l1</i> RV      | TGCTGGTAGAACACATTGGAG        |
| <i>Npc1l1</i> Probe   | AGCTGAACTACGGAAGGTGCCTG      |
| <i>Fabp1</i> FW       | TCTCCGGCAAGTACCAATTG         |
| <i>Fabp1</i> RV       | TTGATGTCCTTCCCTTTCTGG        |
| <i>Fabp1</i> Probe    | TGAATGGCTCAAAGTTCTCCTGGCT    |
| <i>Fabp2</i> FW       | AGCTCGGTGTAACTTTCCC          |
| <i>Fabp2</i> RV       | TTCATTACCAGAAACCTCTCGG       |
| <i>Fabp2</i> Probe    | TTATTTCCCTCAATGGTCCAGGCC     |
| <i>Il12p40</i> FW     | GGCTGGTGCAAAGAAACATGGACT     |
| <i>Il12p40</i> RV     | AGAGACGCCATTCCACATGTCACT     |
| <i>Il12p40</i> Probe  | TCAACATCAAGAGCAGTAGCAGTTCCC  |

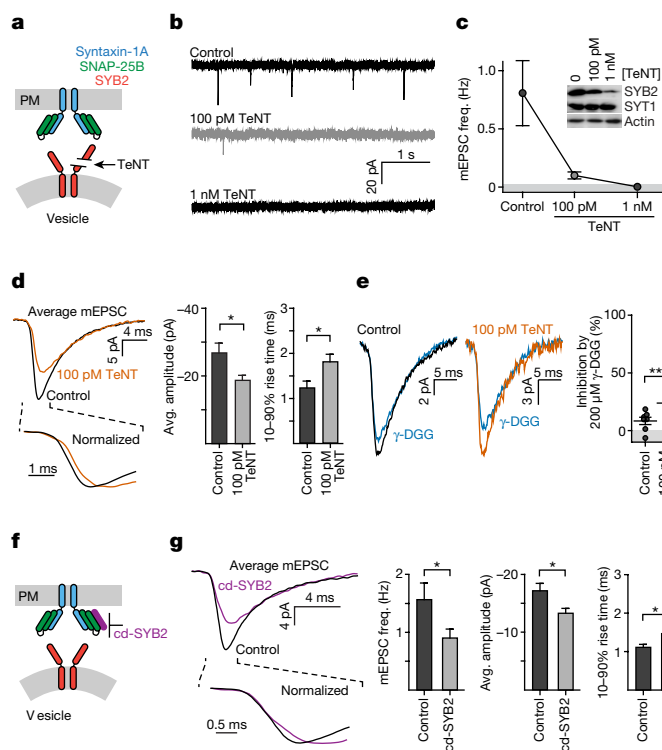
# Dynamics and number of *trans*-SNARE complexes determine nascent fusion pore properties

Huan Bao<sup>1,2\*</sup>, Debasis Das<sup>1,2\*</sup>, Nicholas A. Courtney<sup>1,2</sup>, Yihao Jiang<sup>1</sup>, Joseph S. Briguglio<sup>1,2</sup>, Xiaochu Lou<sup>1,2</sup>, Daniel Roston<sup>3</sup>, Qiang Cui<sup>3</sup>, Baron Chanda<sup>1,4</sup> & Edwin R. Chapman<sup>1,2</sup>

The fusion pore is the first crucial intermediate formed during exocytosis, yet little is known about the mechanisms that determine the size and kinetic properties of these transient structures<sup>1</sup>. Here, we reduced the number of available SNAREs (proteins that mediate vesicle fusion) in neurons and observed changes in transmitter release that are suggestive of alterations in fusion pores. To investigate these changes, we employed reconstituted fusion assays using nanodiscs to trap pores in their initial open state. Optical measurements revealed that increasing the number of SNARE complexes enhanced the rate of release from single pores and enabled the escape of larger cargoes. To determine whether this effect was due to changes in nascent pore size or to changes in stability, we developed an approach that uses nanodiscs and planar lipid bilayer electrophysiology to afford microsecond resolution at the single event level. Both pore size and stability were affected by SNARE copy number. Increasing the number of vesicle (v)-SNAREs per nanodisc from three to five caused a twofold increase in pore size and decreased the rate of pore closure by more than three orders of magnitude. Moreover, pairing of v-SNAREs and target (t)-SNAREs

to form *trans*-SNARE complexes was highly dynamic: flickering nascent pores closed upon addition of a v-SNARE fragment, revealing that the fully assembled, stable SNARE complex does not form at this stage of exocytosis. Finally, a deletion at the base of the SNARE complex, which mimics the action of botulinum neurotoxin A, markedly reduced fusion pore stability. In summary, *trans*-SNARE complexes are dynamic, and the number of SNAREs recruited to drive fusion determines fundamental properties of individual pores.

To understand how membranes fuse during exocytosis, the structure and dynamics of the first crucial intermediate, the fusion pore, must be determined<sup>1</sup>. Moreover, fusion pore properties can affect cargo release from neuroendocrine cells<sup>2</sup>, and can potentially alter aspects of synaptic transmission<sup>3,4</sup>. For example, small unstable pores would allow only transient release of small hormones from neuroendocrine cells, and could, in principle, limit the rate of glutamate release from nerve terminals to reduce postsynaptic responses during synaptic transmission<sup>3,5,6</sup>. Surprisingly little is known about the factors that determine the size and dynamics of fusion pores, but it has been hypothesized that at

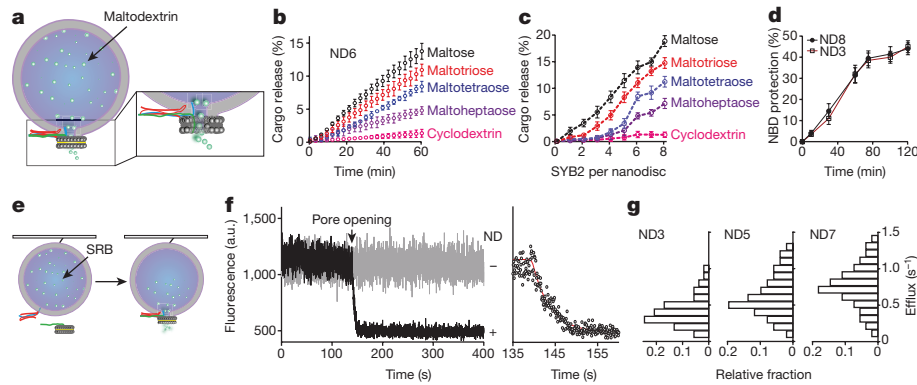


**Figure 1 | Reducing v-SNARE or t-SNARE availability alters the shape of mEPSCs. a**, Cleavage of SYB2 by TeNT. PM, plasma membrane.

**b, c**, Representative traces (**b**) and quantification of mEPSC frequency (freq.; **c**) after treatment with TeNT. Control:  $0.81 \pm 0.28$  Hz (95% confidence interval 0.18–1.44),  $n = 10$ ; 100 pM TeNT:  $0.10 \pm 0.02$  Hz (0.04–0.17),  $n = 10$ ;  $P < 0.001$  for comparison of all three groups by Kruskal–Wallis  $U$ -test;  $P = 0.001$  for control versus 100 pM, Dunn's multiple comparison post hoc test. Inset in **c** shows immunoblot of SYB2,  $\beta$ -actin and synaptotagmin 1 (SYT1) in control and TeNT-treated neurons. Similar results were obtained in three independent trials. **d**, Averaged mEPSC traces after treatment with TeNT (left). Amplitudes, control:  $-27 \pm 3$  pA ( $-20$  to  $-33$ ),  $n = 10$  neurons; 100 pM TeNT:  $-19 \pm 2$  pA ( $-15$  to  $-22$ ),  $n = 10$  neurons;  $P = 0.021$ , two-tailed  $t$ -test. Ten to ninety per cent rise times, control:  $1.2 \pm 0.2$  ms ( $0.9$ – $1.6$  ms),  $n = 10$  neurons; 100 pM TeNT:  $1.8 \pm 0.2$  ms ( $1.3$ – $2.2$ ),  $n = 10$  neurons;  $P = 0.020$ , two-tailed  $t$ -test. **e**, Averaged traces (left) and quantification (right) of  $\gamma$ -DGG-mediated inhibition of mEPSCs. Control:  $8 \pm 3\%$  reduction in amplitude ( $1$ – $16$ ),  $n = 7$  neurons; 100 pM TeNT:  $23 \pm 3\%$  inhibition ( $17$ – $31$ ),  $n = 6$  neurons;  $P = 0.005$ , two-tailed Mann–Whitney test. **a–e**, Experiments were performed using one coverslip from each of three independent litters of mice. **f**, Cd-SYB2 occupies t-SNAREs to inhibit fusion. **g**, Averaged mEPSC traces (left); frequencies, amplitudes, and rise times are plotted on the right. Frequency: control:  $1.6 \pm 0.3$  Hz ( $1.0$ – $2.1$ ),  $n = 19$ ; cd-SYB2:  $0.9 \pm 0.2$  Hz ( $0.6$ – $1.2$ ),  $n = 21$ ;  $P = 0.049$ , two-tailed Mann–Whitney test. Amplitude: control:  $17 \pm 1$  pA ( $-15$  to  $-20$ ); cd-SYB2:  $-13 \pm 1$  pA ( $-12$  to  $-15$ );  $P = 0.017$ , Welch's two-tailed  $t$ -test. Rise time: control:  $1.1 \pm 0.1$  ms ( $0.9$ – $1.3$ ); cd-SYB2:  $1.4 \pm 0.1$  ms ( $1.2$ – $1.7$ );  $P = 0.021$ , Welch's two-tailed  $t$ -test. In **f, g**, experiments were performed using two litters, two coverslips per litter. \* $P < 0.05$ ; \*\* $P < 0.01$ . Data are presented as mean  $\pm$  s.e.m. (95% confidence interval).

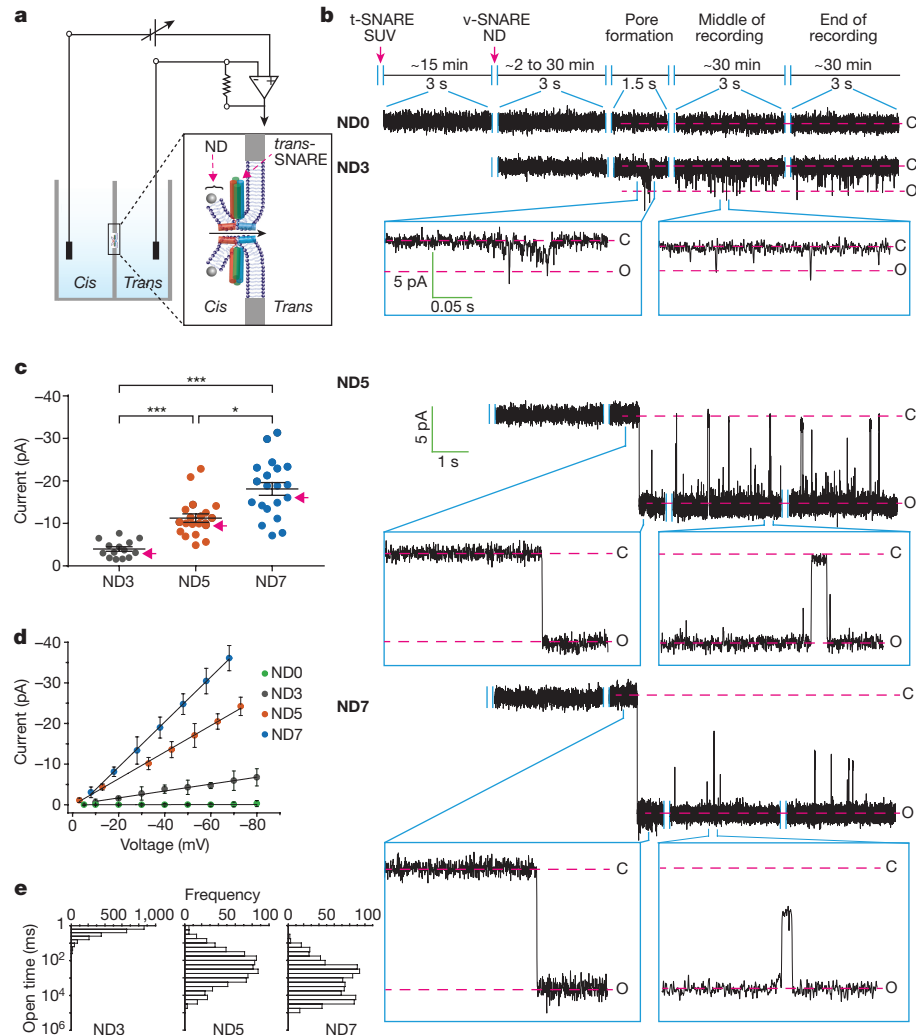
<sup>1</sup>Department of Neuroscience, University of Wisconsin–Madison, 1111 Highland Avenue, Madison, Wisconsin 53705, USA. <sup>2</sup>Howard Hughes Medical Institute, 1111 Highland Avenue, Madison, Wisconsin 53705, USA. <sup>3</sup>Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>4</sup>Department of Biomolecular Chemistry, University of Wisconsin, Madison, 420 Henry Mall, Madison, Wisconsin 53706, USA.

\*These authors contributed equally to this work.



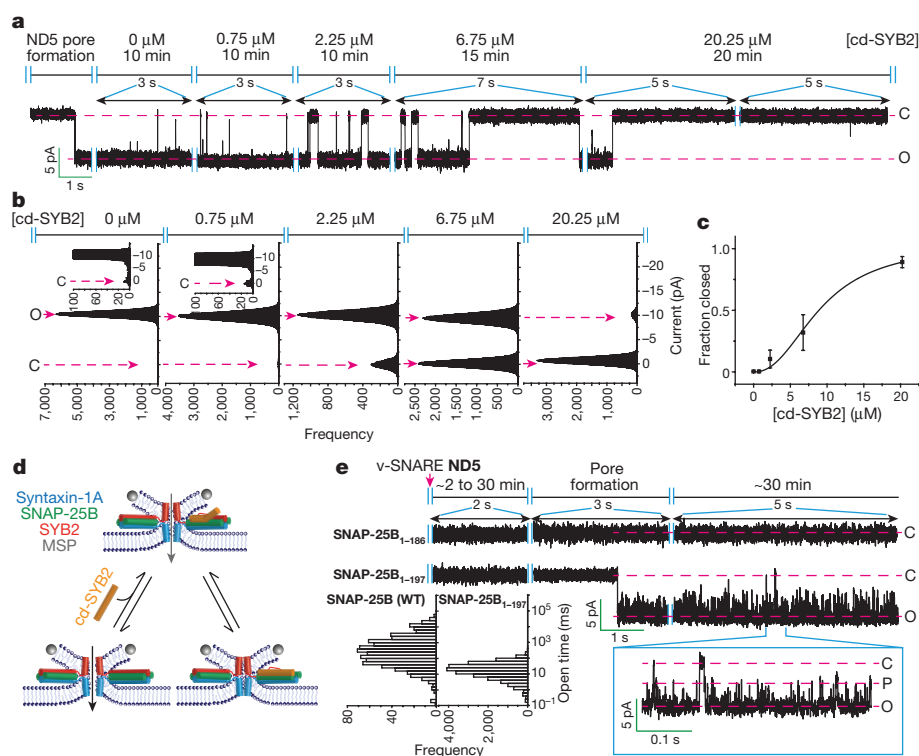
**Figure 2 | Reconstituted fusion assays reveal changes in cargo efflux rates as a function of SYB2 copy number.** **a**, Ensemble fusion assay using nanodiscs and SUVs **b**, Release time courses for different maltodextrins from vesicles using ND6. **c**, Maltodextrin release efficiency against SYB2 copy number per nanodisc. **d**, Dithionite quenching of *N*-(7-nitro-2-1,3-benzoxadiazol-4-yl) (NBD)-labelled lipid revealed that the number of open pores was the same for ND3 and ND8. Data in **b–d** are shown

as mean  $\pm$  s.d.,  $n = 3$  independent experiments. **e**, Illustration of single-vesicle fusion assays. **f**, Left, representative trace showing sulforhodamine b (SRB) efflux through a single fusion pore with or without nanodiscs (ND); right, expanded time scale, fitted with a single exponential function (red). Similar results were obtained in four independent experiments. a.u., arbitrary units. **g**, SRB efflux rates using ND3, ND5 and ND7;  $n = 54, 51$ , and 53 respectively; four independent trials per condition.



**Figure 3 | Properties of single fusion pores measured by planar lipid bilayer electrophysiology.** **a**, Illustration of the nanodisc-BLM assay. **b**, Traces of single pores at  $\Delta\Psi = -50$  mV for ND0, ND3, ND5 and ND7. Closed (C) and open (O) states are indicated, along with the respective currents. **c**, Pore currents obtained using the indicated nanodisc; red arrows indicate representative pores shown in **b**. ND3:  $-4 \pm 1$  pA ( $-3$  to  $-5$  pA); ND5:  $-11 \pm 1$  pA ( $-9$  to  $-13$  pA); ND7:  $-18 \pm 2$  pA ( $-15$  to  $-21$  pA);  $P < 0.001$ , Kruskal-Wallis *U*-test. ND3 versus ND5,

$P < 0.001$ ; ND3 versus ND7,  $P < 0.001$ ; ND5 versus ND7,  $P = 0.036$ , Dunn's multiple comparison post hoc tests. Data shown as mean  $\pm$  s.e.m. (95% confidence interval). **d**, Current-voltage (*I*-*V*) relationships for pores formed using the indicated nanodiscs. Data are presented as mean  $\pm$  s.e.m. **e**, Open dwell-time histograms of pores.  $n = 14, 20$  and 20 independent BLMs for ND3, ND5 and ND7 respectively; five sets of nanodiscs of each type were used. \*\*\* $P < 0.001$ ; \* $P = 0.05$ .



**Figure 4 | Trans-SNARE complexes are dynamic.** **a, b**, Representative recording (**a**) and current histogram (**b**) of a pore formed using ND5, before and after addition of cd-SYB2 at the indicated concentrations. **c**, Fraction of time for which pores were fully closed, plotted as a function of [cd-SYB2].  $n = 11$  independent BLMs; five sets of nanodiscs. Data are presented as mean  $\pm$  s.e.m. **d**, Illustration of dynamic *trans*-SNARE pairing, as evidenced by cd-SYB2-mediated destabilization of open pores. **e**, Open dwell time histograms for nanodisc-BLM assays performed using ND5 and the indicated SNAP-25B mutants. SNAP-25B<sub>1-197</sub> gave rise to partial conductances, denoted P.  $n = 5$  independent BLMs for each mutant; three sets of nanodiscs.

least three SNARE complexes are required to hold fusion pores open and allow efficient cargo release in a reconstituted system<sup>7</sup>. Here, we used a combination of approaches to determine directly whether the number of SNAREs that drive fusion dictate fundamental properties of nascent fusion pores, and whether *trans*-SNARE complexes are stable or dynamic.

To examine how SNARE copy number influences neurotransmitter efflux from synaptic vesicles, we recorded  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor (AMPA)-mediated miniature excitatory postsynaptic currents (mEPSCs) in neuronal cultures treated with tetanus toxin (TeNT) (Fig. 1a–c). Although higher doses nearly abolished mEPSCs (Fig. 1b, c), 100 pM TeNT, which caused 50% cleavage of the vesicle-associated membrane protein SYB2 (also known as VAMP2; Fig. 1c), yielded a sufficient frequency of mEPSCs to enable quantitative analysis (Fig. 1b). Notably, mEPSCs that remained after TeNT treatment had smaller amplitudes and slower rise kinetics (Fig. 1d) than untreated controls, consistent with a previous report examining miniature *N*-methyl-D-aspartate (NMDA)-mediated events<sup>6</sup>. Application of the rapidly dissociating, competitive AMPAR antagonist  $\gamma$ -D-glutamyl-glycine ( $\gamma$ -DGG, 200  $\mu$ M) inhibited mEPSCs in TeNT-treated neurons more potently than in untreated control cells (Fig. 1e). As inhibition by  $\gamma$ -DGG is inversely proportional to the glutamate concentration at receptors<sup>8</sup>, these findings suggest that TeNT treatment impaired the ability of glutamate to escape synaptic vesicles into the synaptic cleft. We tested this hypothesis further by overexpressing the cytosolic domain of SYB2 (cd-SYB2), which binds native t-SNAREs to inhibit *trans*-SNARE complex formation (Fig. 1f, g, Extended Data Fig. 1). Neuronal cultures expressing cd-SYB2 had fewer mEPSCs than control neurons, and these were smaller in amplitude and slower to rise. Together, these experiments suggest that the abundance of *trans*-SNARE pairs might modulate the efflux of glutamate through fusion pores.

To test this hypothesis directly, we devised an *in vitro* assay to probe pores using cargoes of different sizes. As shown previously, fusion between small unilamellar vesicles (SUVs) and 13-nm nanodiscs results in pores that cannot dilate, owing to the rigid framework of the nanodisc, thus enabling the biochemical characterization of pores in their initial open state<sup>7,9</sup>. We encapsulated a variety of maltodextrins

in t-SNARE SUVs and incubated them with v-SNARE-bearing nanodiscs that harboured one to eight copies of SYB2, designated ND1–ND8. Flux of cargo through fusion pores was monitored using an optical sensor that recognizes each of the maltodextrins used (Fig. 2a–c, Extended Data Fig. 2a, b). Using a fixed number of SNAREs (ND6), the rank order of cargo release rates correlated with cargo size (Fig. 2b): the smaller the cargo, the more quickly it escaped through fusion pores. As the number of SNAREs per fusion reaction was increased from one to eight, the rate of efflux of each maltodextrin (except for cyclodextrin) also increased, and larger cargoes were able to escape (Fig. 2c; estimates of pore size are provided in Extended Data Fig. 2c, d). Using a dithionite quenching assay, we found that the number of SNAREs did not influence the total number of open fusion pores that were formed (Fig. 2d, Extended Data Fig. 2e). Together, these data show that SNARE copy number determines the size and/or kinetic stability of individual fusion pores; these findings were confirmed using single-vesicle fusion assays (Fig. 2e–g, Extended Data Fig. 2f, g).

To achieve sub-millisecond time resolution, we developed an approach for monitoring recombinant fusion pores electrophysiologically. t-SNAREs were reconstituted into black lipid membranes (BLMs), at a density of 0.4 molecules per  $\mu$ m<sup>2</sup> (Extended Data Fig. 3c, d) in a planar lipid bilayer electrophysiology setup<sup>10</sup> (Fig. 3, Extended Data Figs 3a, b, e and 4). Addition of v-SNARE-bearing nanodiscs into the *cis* chamber resulted in the formation of single fusion pores, as evidenced by the currents detected (Fig. 3a, b). Control experiments establish that these are bona fide fusion pores that result from *trans*-SNARE pairing (Extended Data Table 1). Remarkable differences were observed among pores formed by ND3, ND5 and ND7 (Fig. 3b–e). At  $-50$  mV, ND3 produced pores that remained closed most of the time and flickered open only transiently (Fig. 3b, ND3). By contrast, pores formed by ND5 remained open most of the time, but closed transiently (Fig. 3b, ND5). Finally, ND7 pores remained open during the entire recording period; these pores flickered but never closed completely (Fig. 3b, ND7). To estimate pore size, we generated *I*–*V* plots (Fig. 3d). From the conductance values, the estimated diameters of pores formed using ND3, 5 and 7 were  $1.1 \pm 0.3$  nm,  $2.2 \pm 0.3$  nm and  $2.9 \pm 0.3$  nm, respectively, consistent with the range of pore sizes observed using nanodiscs

and 'flipped' t-SNAREs on the surface of cells<sup>11</sup>. We made similar observations using 50-nm nanodiscs (Extended Data Fig. 5).

Kinetic analysis revealed differences in the open dwell-time distribution between ND3 and ND5 or ND7 (Fig. 3e): increasing the number of SNAREs notably enhanced the stability of the open state (Extended Data Table 2). Moreover, even though the kinetic stabilities of pores formed using ND5 and ND7 were similar, pore size still increased at the higher copy number. Therefore, the size and dynamics of individual pores are differentially regulated by SNARE copy number. Similar results were obtained using yeast SNAREs<sup>12</sup>, establishing the generality of these findings (Extended Data Fig. 6, Extended Data Table 2).

As the nanodisc–BLM recordings showed that recombinant fusion pores, under all the measured conditions, rapidly convert between open and at least partially closed states, we hypothesized that the underlying *trans*-SNARE complexes exist in metastable conformational states. To test this hypothesis, we titrated cd-SYB2 onto pre-formed pores assembled using ND5 (Fig. 4a–d). Pores initially destabilized, and at the highest dose all pores eventually closed (Fig. 4a–c); at lower doses of cd-SYB2, we sometimes observed partial closure (Extended Data Fig. 7a, b). Addition of cd-SYB2 also closed fusion pores formed using ND7, albeit with reduced potency (Extended Data Fig. 7c, d). In control experiments, bovine serum albumin had no effect on pores and cd-SYB2(4A), a mutant with impaired t-SNARE binding activity<sup>13</sup>, had only limited effects at the highest dose tested (Extended Data Fig. 8a, b). Together, these findings demonstrate that in contrast to *cis*-SNARE complexes, which are highly stable<sup>14</sup>, *trans*-SNARE interactions are dynamic and potentially reversible, even after pores have opened (Fig. 4d). Consistent with this conclusion, impairment of *trans*-SNARE interactions via a C-terminal truncation of SNAP-25B (consisting of residues 1–197) that mimics cleavage by botulinum neurotoxin A results in a marked increase in flickering behaviour without affecting pore size (Fig. 4e, Extended Data Fig. 8c). Truncation of twenty residues (to leave residues 1–186) completely abolished pore formation (Fig. 4e). These data indicate that *trans*-SNARE interactions, at the base of the SNARE complex, control pore dynamics.

In summary, the exocytotic fusion pore corresponds to the initial, narrow channel formed between secretory vesicles and the plasma membrane. Release of neurotransmitters and hormones occurs via diffusion through this transient structure before, or even without, dilation<sup>3</sup>. Previous electrophysiological measurements revealed a range of pore sizes<sup>15,16</sup>, as well as flickering behaviour<sup>17,18</sup>, in cells; these observations are recapitulated in the nanodisc–BLM system described here. Our results provide direct experimental support for the idea that a certain number of SNAREs is needed to hold fusion pores open<sup>7</sup>, with more SNAREs resulting in larger fusion pores<sup>19</sup>. Moreover, even after fusion pores have opened, *trans*-SNARE complexes remain dynamic and reversible. It will be interesting to determine how far a fusion pore must dilate in order for the SNARE complex to become irreversible, and to ascertain the effect of myriad regulatory factors on the properties of individual pores<sup>20,21</sup>. Finally, it will also be interesting to determine whether the findings concerning SNARE copy number reported here apply to other cellular fusogens<sup>22</sup>, including atlastin (homotypic endoplasmic reticulum fusion<sup>23</sup>), mitofusin 1 and 2 (mitochondria fusion<sup>24</sup>), and the proteins that mediate ectoplasmic fusion<sup>25</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 March; accepted 20 December 2017.**

**Published online 31 January 2018.**

- Giraudo, C. G. *et al.* SNAREs can promote complete fusion and hemifusion as alternative outcomes. *J. Cell Biol.* **170**, 249–260 (2005).
- Fulop, T., Radabaugh, S. & Smith, C. Activity-dependent differential transmitter release in mouse adrenal chromaffin cells. *J. Neurosci.* **25**, 7324–7332 (2005).

- Richards, D. A. Vesicular release mode shapes the postsynaptic response at hippocampal synapses. *J. Physiol. (Lond.)* **587**, 5073–5080 (2009).
- Choi, S., Klingauf, J. & Tsien, R. W. Fusion pore modulation as a presynaptic mechanism contributing to expression of long-term potentiation. *Phil. Trans. R. Soc. Lond. B* **358**, 695–705 (2003).
- Alabi, A. A. & Tsien, R. W. Perspectives on kiss-and-run: role in exocytosis, endocytosis, and neurotransmission. *Annu. Rev. Physiol.* **75**, 393–422 (2013).
- Renger, J. J., Egles, C. & Liu, G. A developmental switch in neurotransmitter flux enhances synaptic efficacy by affecting AMPA receptor activation. *Neuron* **29**, 469–484 (2001).
- Shi, L. *et al.* SNARE proteins: one to fuse and three to keep the nascent fusion pore open. *Science* **335**, 1355–1359 (2012).
- Liu, G., Choi, S. & Tsien, R. W. Variability of neurotransmitter concentration and nonsaturation of postsynaptic AMPA receptors at synapses in hippocampal cultures and slices. *Neuron* **22**, 395–409 (1999).
- Bao, H. *et al.* Exocytotic fusion pores are composed of both lipids and proteins. *Nat. Struct. Mol. Biol.* **23**, 67–73 (2016).
- Mueller, P., Rudin, D. O., Tien, H. T. & Wescott, W. C. Reconstitution of cell membrane structure *in vitro* and its transformation into an excitable system. *Nature* **194**, 979–980 (1962).
- Wu, Z. *et al.* Nanodisc–cell fusion: control of fusion pore nucleation and lifetimes by SNARE protein transmembrane domains. *Sci. Rep.* **6**, 27287 (2016).
- Bhalla, A., Chicka, M. C., Tucker, W. C. & Chapman, E. R. Ca<sup>2+</sup>-synaptotagmin directly regulates t-SNARE function during reconstituted membrane fusion. *Nat. Struct. Mol. Biol.* **13**, 323–330 (2006).
- Wiederhold, K. *et al.* A coiled coil trigger site is essential for rapid binding of synaptobrevin to the SNARE acceptor complex. *J. Biol. Chem.* **285**, 21549–21559 (2010).
- Südhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323**, 474–477 (2009).
- Albillos, A. *et al.* The exocytotic event in chromaffin cells revealed by patch amperometry. *Nature* **389**, 509–512 (1997).
- Klyachko, V. A. & Jackson, M. B. Capacitance steps and fusion pores of small and large-dense-core vesicles in nerve terminals. *Nature* **418**, 89–92 (2002).
- Henkel, A. W., Meiri, H., Horstmann, H., Lindau, M. & Almers, W. Rhythmic opening and closing of vesicles during constitutive exo- and endocytosis in chromaffin cells. *EMBO J.* **19**, 84–93 (2000).
- Staal, R. G. W., Mosharov, E. V. & Sulzer, D. Dopamine neurons release transmitter via a flickering fusion pore. *Nat. Neurosci.* **7**, 341–346 (2004).
- Wu, Z. Y. *et al.* Dilation of fusion pores by crowding of SNARE proteins. *Elife* **6**, e22964 (2017).
- Sinha, R., Ahmed, S., Jahn, R. & Klingauf, J. Two synaptobrevin molecules are sufficient for vesicle fusion in central nervous system synapses. *Proc. Natl Acad. Sci. USA* **108**, 14318–14323 (2011).
- Lai, Y. *et al.* Fusion pore formation and expansion induced by Ca<sup>2+</sup> and synaptotagmin 1. *Proc. Natl Acad. Sci. USA* **110**, 1333–1338 (2013).
- Wickner, W. & Schekman, R. Membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 658–664 (2008).
- Orso, G. *et al.* Homotypic fusion of ER membranes requires the dynamin-like GTPase atlastin. *Nature* **460**, 978–983 (2009).
- Cao, Y. L. *et al.* MFN1 structures reveal nucleotide-triggered dimerization critical for mitochondrial fusion. *Nature* **542**, 372–376 (2017).
- Podbilewicz, B. Virus and cell fusion mechanisms. *Annu. Rev. Cell Dev. Biol.* **30**, 111–139 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This study was supported by grants from the NIH (MH061876 and NS097362 to E.R.C.; NS081293 to B.C.). H.B. was supported by a postdoctoral fellowship from the Human Frontier Science Program. D.R. was supported by an NIH fellowship (F32GM112371). E.R.C. is an Investigator of the Howard Hughes Medical Institute.

**Author Contributions** H.B. and E.R.C. conceived of the project and designed the biochemistry experiments. H.B. performed nanodisc reconstitution and fusion assays. H.B. and D.D. performed the planar lipid bilayer recordings. N.A.C. designed and conducted the experiments using neurons. Y.J. and B.C. aided in the initial planar lipid bilayer recordings. J.S.B. contributed neurons. X.L. and H.B. contributed to the single vesicle fusion assays. D.R. and Q.C. conducted molecular dynamics simulations. H.B., D.D., N.A.C. and E.R.C. wrote the paper, and all other authors edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.R.C. ([chapman@wisc.edu](mailto:chapman@wisc.edu)).

**Reviewer Information** Nature thanks J. Dittman, R. Heidelberger, J. Sørensen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Reagents.** 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(biotinyl) (biotin-PE), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (PC), 1,2-dioleoyl-*sn*-glycero-3-phospho-L-serine (PS), 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (PE), 2-dioleoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol) (PG), 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(7-nitro-2-1,3-benzoxadiazol-4-yl) (NBD-PE) and 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(lissamine rhodamine B sulfonyl) (rhodamine-PE) were purchased from Avanti Polar Lipids.  $\gamma$ -DGG was obtained from Abcam. All other chemicals were from Sigma-Aldrich.

**Cell culture and lentivirus.** Cultured rat cortical neurons were prepared from embryonic day 18–19 Sprague Dawley rats as previously described<sup>26</sup>. Sex was not determined, and neurons from all of the pups in each litter were pooled during culturing. In brief, neurons were plated on poly-D-lysine-coated glass coverslips (12 mm) at a density of 100,000 cm<sup>-2</sup>. Neurons were cultured in Neurobasal A medium (Gibco) supplemented with B-27 (2%, Gibco) and GlutaMAX (2 mM, Gibco) and maintained at 37 °C in a 5% CO<sub>2</sub> humidified incubator. After day 13 or 14 *in vitro*, half of the neuronal coverslips were treated with tetanus toxin (1 nM or 100 pM, List Biological Labs) at 37 °C for 24 h. Untreated neurons were used as controls. Electrophysiological recordings of both treated and untreated neurons were performed immediately following the 24-h treatment period (day 14–15 *in vitro*). All procedures were approved by the Animal Care and Use Committee at the University of Wisconsin and performed in accordance with the guidelines of the National Institutes of Health.

For the viral expression experiments, DNA sequences encoding either cd-SYB2 (residues 1–95 of SYB2) or GFP were subcloned into a FUGW transfer plasmid modified with a synapsin promoter and an IRES-expressed soluble GFP marker. Lentivirus particles were generated by co-transfection of the transfer plasmid and helper plasmids (pCD/NL-BH\* $\Delta\Delta\Delta$  and VSV-G encoding pLTR-G) into HEK293T/17 cells (ATCC, not tested for mycoplasma contamination)<sup>27</sup>. The supernatant was collected after 48–72 h of expression, filtered through a 0.45- $\mu$ m PVDF filter, and concentrated by ultra-centrifugation at 110,000g for 2 h. Viral particles were resuspended in Ca<sup>2+</sup>/Mg<sup>2+</sup>-free PBS and used to infect neurons at day 6 *in vitro*. Electrophysiological recordings were then performed at day 14–15 *in vitro*.

**Immunocytochemistry.** At 14 days *in vitro*, cell cultures were fixed for 15 min with 4% paraformaldehyde (wt/vol) in PBS, permeabilized for 10 min with 0.2% saponin (wt/vol), and blocked for 60 min with 10% goat serum (vol/vol, Abcam) plus 0.1% Tween-20 (vol/vol). Coverslips were then incubated with primary antibodies (anti-GFP: Abcam, 1:1,000, chicken; anti-MAP2: EMD, 1:1,000, mouse) at room temperature for 1 h. Samples were washed three times with 0.02% saponin (wt/vol) in PBS and labelled with Alexa Fluor 488-tagged anti-chicken and Alexa Fluor 546-tagged anti-mouse IgG (1:400, Invitrogen) for 1 h at room temperature. Samples were again washed three times and mounted in Fluoromount G mounting medium (Southern Biotech). Images were obtained using an FV1000 laser-scanning confocal microscope (Olympus) with FV10-ASW 3.1 acquisition software, using a 20 $\times$  1.0 NA water objective, under identical laser and gain settings. Images were analysed using ImageJ (NIH).

**Protein purification and reconstitution.** Membrane scaffold protein (MSP) for 13 nm<sup>9</sup> and 50 nm<sup>28</sup> nanodiscs, the maltose sensor<sup>29</sup>, neuronal (rat SYB2, syntaxin-1A and SNAP-25B) and yeast (Snc2p, Sso1p and Sec9c (residues 401–651)) SNARE proteins were purified as described previously<sup>12</sup>. t-SNARE complexes bearing truncated SNAP-25B (corresponding to residues 1–197 or residues 1–186) were also prepared and studied; the former truncation mimics cleavage by botulinum neurotoxin A<sup>30</sup>. To prepare t-SNARE vesicles, lipids (10% PE, 15% PS and 75% PC) and the t-SNARE heterodimer were incubated with the respective cargoes and 2% octyl- $\beta$ -glucoside on ice for 30 min. Detergent was removed by addition of Biobeads (Bio-Rad) (one-third volume) followed by gentle shaking (at 4 °C, overnight). The mixture was extruded through a 0.2- $\mu$ m filter and the t-SNARE vesicles were purified by passing through a PD10 column (5 ml) equilibrated in reconstitution buffer (25 mM HEPES, pH 7.5, 100 mM KCl, 1 mM DTT). Finally, purified t-SNARE vesicles were dialysed against reconstitution buffer (4 °C, overnight). Reconstitution of SYB2 into 13-nm nanodiscs was performed as described<sup>9</sup>. For reconstitution of SYB2 into 50-nm nanodiscs, the MSP:lipid ratio was 2:4:000. To incorporate different copy numbers of SYB2 into 50-nm nanodiscs, the following MSP:SYB2 ratios were used: 2:2 (ND3), 2:4 (ND5) and 2:10 (ND7). The reconstituted nanodiscs were incubated with Ni<sup>2+</sup>-NTA resin to remove SYB2-free nanodiscs. Nanodiscs containing SYB2 were eluted with reconstitution buffer containing 0.4 M imidazole. The nanodiscs were further purified via sucrose density-gradient centrifugation<sup>31</sup>, followed by dialysis against reconstitution buffer (4 °C,

overnight). The copy number of SYB2 per nanodisc refers to the total number of SYB2 molecules, not the number of copies per face of the nanodisc.

**Ensemble fusion assays.** Maltodextrin release assays were carried out using the maltose sensor (1  $\mu$ M)<sup>29</sup>, SYB2 nanodiscs (0.2  $\mu$ M), and t-SNARE vesicles (1  $\mu$ M) containing maltodextrins, at 37 °C in reconstitution buffer. The fluorescence of the sensor was monitored for 1 h using a plate reader (HT synergy, BioTek). After each run, 1  $\mu$ M melittin was added to each sample, and data were collected for another 30 min. Melittin forms channels to release all the maltodextrin from each vesicle, thus producing the maximum fluorescence signal (100%) that can be obtained. Data were collected from three independent experiments.

Efflux rates were used to estimate fusion pore size. As described previously<sup>7</sup>, the time it takes for a single cargo molecule to traverse the fusion pore is  $\frac{4R^2}{\alpha(r_p - r_c)^2}$ ,

in which  $\alpha$  is the frequency of collisions of the cargo molecule with the membrane that forms the t-SNARE vesicle, and  $R$ ,  $r_p$  and  $r_c$  are the radii of the liposome, fusion pore and cargo, respectively. Thus, the difference in the release rates between maltose and maltotetraose is given by:

$$\frac{K_{\text{maltose}}}{K_{\text{maltotetraose}}} = \frac{(r_p - r_{\text{maltose}})^2}{(r_p - r_{\text{maltotetraose}})^2},$$

in which  $K_{\text{maltose}}$  and  $K_{\text{maltotetraose}}$  are the respective release rates for these sugars, and  $r_{\text{maltose}}$  (0.35 nm) and  $r_{\text{maltotetraose}}$  (0.42 nm) are their radii as determined by molecular dynamics simulations.

**Molecular dynamics simulations to estimate maltodextrin size.** All calculations used the program CHARMM<sup>32</sup>, and dynamics calculations were conducted through the CHARMM interface with OpenMM<sup>33</sup>. The sugars were treated with the CHARMM glycan force field<sup>34,35</sup> and water was treated with the TIP3P model<sup>36</sup>. Initial structures for the sugars were obtained from the online tool SWEET<sup>37</sup>. The online tool CHARMM-GUI<sup>38,39</sup> was used to construct the initial setup for each sugar. The sugars were each dissolved in a cubic box of water such that the nearest edge of the box was at least 10 Å away. After a brief geometry optimization, the systems were heated to 298 K (150 ps), followed by a 1-ns equilibration at 298 K and 1 atm using periodic boundary conditions. The simulations employed the NPT ensemble using the Andersen thermostat and MC barostat. Non-bonded interactions were cut off above a distance of 12 Å with a switching function from 10 Å to 12 Å, and the integration time step was 1 fs. Following equilibration, 10-ns production runs were used to determine the principal axes shown in Extended Data Fig. 2.

**Single vesicle fusion assays.** A prism-based total internal reflection (TIRF) microscopy setup and associated flow chambers were prepared as described previously<sup>40</sup>. t-SNARE vesicles containing SRB were prepared and immobilized on the surface of quartz slides as described<sup>40</sup>. The trapping efficiency of SRB in vesicles was ~2% of the total SRB and the [SRB] was ~1 mM per vesicle. SYB2 nanodiscs (100  $\mu$ l, 50 nM) were injected at the indicated time for 10 s and data were recorded for an additional 400 s. Leakage and photobleaching of SRB were negligible (<7% fluorescence decrease). By contrast, opening of fusion pores led to fluorescence decreases of >50%. Fusion probability was defined as the fraction of tethered SUVs in which a pore opening event was observed.

**Planar lipid bilayer electrophysiology.** Planar lipid bilayer recordings that are shown in the main figures were performed using a Planar Lipid Bilayer Workstation (BLM) from Warner Instruments<sup>10</sup> as described<sup>41</sup>. In brief, lipids (75% PE and 25% PG, at 30 mg/ml in *n*-decane) were first painted onto a 150- $\mu$ m aperture in a 1-ml, white Delrin cup (Warner Instruments), allowed to dry for 15 min, and then the aperture was bathed in 1 ml of 25 mM HEPES, pH 7.5 and 100 mM KCl. The lipid solution was gently re-applied to the hole until a conductance-blocking seal was formed, as determined by capacitance measurements. This process was repeated, either with a brush or an air bubble, until the desired capacitance was achieved. Syntaxin-1A/SNAP-25B proteoliposomes (75% PE and 25% PG) were then added to the *cis* chamber of the apparatus; these spontaneously fuse with the planar bilayer, thus depositing the t-SNAREs into the BLM. Then, to form fusion pores, v-SNARE nanodiscs were added to the *cis* chamber. Pores form within 2–30 min and flicker open, or stay open or flicker closed for >90 min. Currents were recorded using Bilayer Clamp Amplifier BC-535 (Warner Instrument) and a Digidata 1550B (with Humsilencer) acquisition system (Molecular Devices). Single-channel recordings were sampled at 10 kHz using pCLAMP10 software (Molecular Devices), and filtered at 5 kHz using a multisection Bessel filter.  $\Delta\psi \equiv \psi_{\text{cis}} - \psi_{\text{trans}}$  ( $\psi_{\text{trans}} \equiv 0$  V). All single channel data were analysed using Clampfit 10.7 (Molecular Devices) and MS Origin 2016 (OriginLab). Histograms of background currents were well fitted by a single Gaussian (centred around 0 pA), whereas current histograms of open fusion pores required a multiple Gaussian model (with the centre of the additional Gaussian representing the mean pore

current). In all figures showing BLM recordings, the representative traces were filtered at 1 kHz for display purposes.

For the experiments reported in Extended Data Table 1, planar lipid bilayer recordings were carried out using the Orbit Mini system (Nanon Technologies). Membranes were painted onto a MECA chip (Nanon Technologies), and fusion pores were formed and analysed as described above, except that data were low-pass filtered at 2 kHz.

The fraction of closed pores (Fig. 4) was calculated using the equation:

$$\text{Fraction closed} = \frac{\text{closed dwell time}}{\text{open dwell time} + \text{closed dwell time}},$$

from a 10-min recording at the indicated [cd-SYB2].

Pore diameters were calculated using the equation  $\frac{1}{\gamma} = \left( l + \frac{\pi r}{2} \right) \frac{\rho}{\pi r^2}$ , in which  $\gamma$  is the pore conductance,  $r$  is the radius,  $l$  is the thickness of the bilayer (10 nm; assuming the pore is a cylinder that spans both the vesicle and target membrane), and  $\rho$  is the resistivity of the buffer (100  $\Omega$  cm)<sup>42</sup>.

**Stopped-flow measurements.** Equal volumes of the maltose sensor (0.1  $\mu$ M) and maltodextrins (indicated concentrations) were mixed in an SX.18MV stopped-flow spectrometer (Applied Photophysics). The samples were excited at 480 nm and the emission was collected at 520 nm using a 10-nm bandpass filter. Data were obtained from three independent experiments.

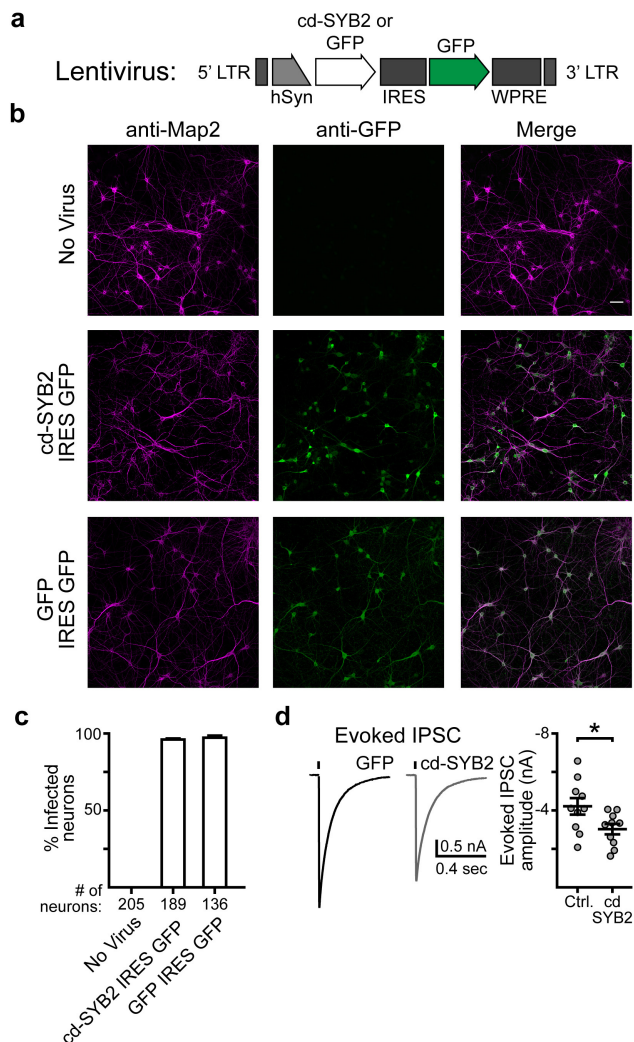
**Electrophysiology.** Whole-cell voltage-clamp recordings were made using a Multiclamp 700B amplifier (Molecular Devices). Recordings were carried out at room temperature in a bath solution containing (in mM): 128 NaCl, 5 KCl, 2 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 30 D-glucose and 25 HEPES, pH 7.3 and 305 mOsm. Patch pipettes (3–5 M $\Omega$ ) were pulled from borosilicate glass (Sutter Instruments). The pipette internal solution contained (in mM): 130 K-gluconate, 1 EGTA, 10 HEPES, 2 ATP, 0.3 GTP, and 5 sodium phosphocreatine, pH 7.35 and 275 mOsm. Data were acquired using a Digidata 1440A (Molecular Devices) and Clampex 10 software (Molecular Devices) at 10 kHz. Neurons were held at –70 mV. Series resistance was compensated and recordings were discarded if the access resistance rose above 15 M $\Omega$  at any point. AMPARs were pharmacologically isolated with 2-amino-5-phosphonovaleate (D-AP5) (50  $\mu$ M, Abcam) and picrotoxin (100  $\mu$ M, Abcam). For mEPSC recordings, tetrodotoxin (TTX, 1  $\mu$ M, Abcam) was included in the bath solution. In some experiments, neurotransmitter release was evoked by a single stimulus using a concentric bipolar electrode (FHC, 125/50  $\mu$ m extended tip). Stimulating electrodes were placed ~100–200  $\mu$ m from the soma being recorded and stimulation currents (0.4–0.7 mA) were adjusted per recording to measure the maximum field-evoked current. For these evoked recordings, the pipette internal solution was modified to include 130 mM KCl (replacing K-gluconate) and 5 mM QX-314 chloride (Tocris) and the bath solution was modified to include CNQX (10  $\mu$ M, Abcam) instead of picrotoxin. Traces were analysed using Clampfit 10 (Molecular Devices).

The two sets of experiments reported in Fig. 1 (a–e versus f and g) were conducted at different times using independent materials, resulting in slightly different values for mEPSC frequencies and amplitudes, with no significant effect on kinetics.

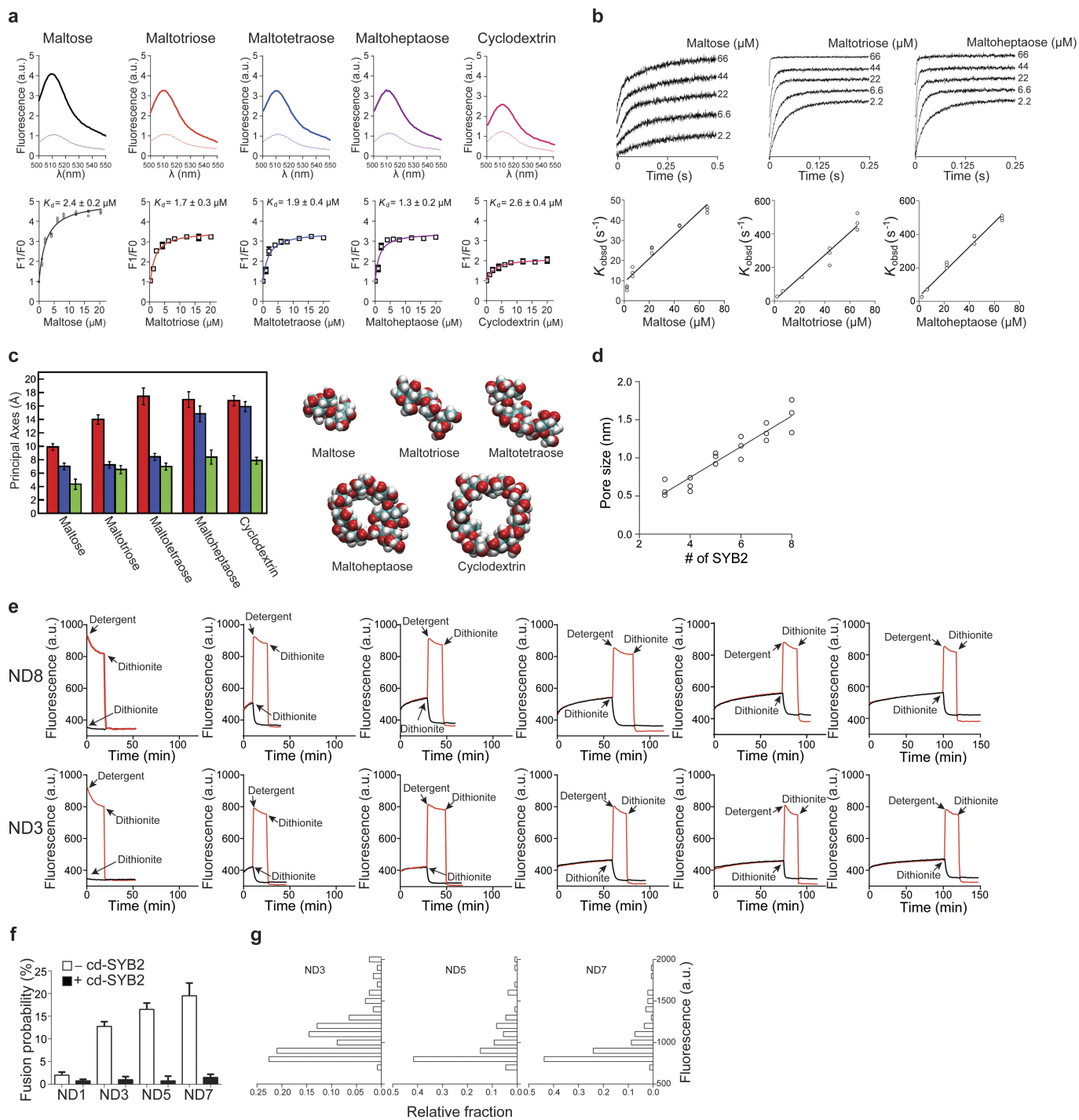
**Other methods.** SDS–PAGE, western blotting, fluorescence spectroscopy, and dithionite quenching assays were performed as described previously<sup>7,9</sup>.

**Data availability.** All original data will be made available by the corresponding authors upon reasonable request. For gel source data, see Supplementary Fig. 1. Source data for Fig. 2b–d are available online.

26. Yeh, F. L. *et al.* SV2 mediates entry of tetanus neurotoxin into central neurons. *PLoS Pathogens* **6**, e1001207 (2010).
27. Kutner, R. H., Zhang, X. Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protocols* **4**, 495–505 (2009).
28. Nasr, M. L. *et al.* Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nat. Methods* **14**, 49–52 (2017).
29. Marvin, J. S., Schreiter, E. R., Echevarria, I. M. & Looger, L. L. A genetically encoded, high-signal-to-noise maltose sensor. *Proteins* **79**, 3025–3036 (2011).
30. Tucker, W. C., Weber, T. & Chapman, E. R. Reconstitution of Ca<sup>2+</sup>-regulated membrane fusion by synaptotagmin and SNAREs. *Science* **304**, 435–438 (2004).
31. Bao, H. & Duong, F. Discovery of an auto-regulation mechanism for the maltose ABC transporter MalFGK2. *PLoS ONE* **7**, e34836 (2012).
32. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
33. Eastman, P. *et al.* OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* **9**, 461–469 (2013).
34. Guvench, O. *et al.* Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564 (2008).
35. Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. & Mackerell, A. D. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353–2370 (2009).
36. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
37. Bohne, A., Lang, E. & von der Lieth, C. W. SWEET—WWW-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics* **15**, 767–768 (1999).
38. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
39. Jo, S., Song, K. C., Desaire, H., MacKerell, A. D. Jr & Im, W. Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **32**, 3135–3141 (2011).
40. Kyoung, M., Zhang, Y., Diao, J., Chu, S. & Brunker, A. T. Studying calcium-triggered vesicle fusion in a single vesicle-vesicle content and lipid-mixing system. *Nat. Protocols* **8**, 1–16 (2013).
41. Finkelstein, A. Bilayers: formation, measurements, and incorporation of components. *Methods Enzymol.* **32**, 489–501 (1974).
42. Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer, 2001).



**Extended Data Figure 1 | Viral expression of cd-SYB2.** **a**, cDNA encoding the cytosolic domain of SYB2 (cd-SYB2, residues 1–95) was cloned into a FUGW transfer vector modified to have a synapsin promoter and to co-express soluble eGFP via an IRES sequence; eGFP serves as a marker for infection efficiency. For control experiments, eGFP alone was expressed. Both constructs were packaged into lentivirus for expression in neuronal cultures. **b**, Representative images of cells stained for a neuronal marker (MAP2, magenta) and GFP (green). Images were adjusted for brightness and contrast for the sake of presentation. Both preparations used for Fig. 1g were examined and had similar GFP expression levels and coverage across cells. The scale bar (50  $\mu\text{m}$ ) applies to all nine images shown. **c**, Quantification of the ICC demonstrating that both cd-SYB2 and control viruses achieved a nearly 100% infection rate. Per cent infected refers to the number of visually identified MAP2-positive somas (that is, neurons) that were also positive for GFP. Three fields of view were quantified for each condition. **d**, Representative traces (left) and quantification (right) demonstrating that cd-SYB2 was expressed at levels sufficient to inhibit evoked IPSCs triggered by field stimulation ( $P = 0.032$ , two-tailed  $t$ -test;  $n = 10$  neurons for each condition, using two litters of mice, three coverslips per condition). Data are presented as mean  $\pm$  s.e.m. \* $P < 0.05$ .

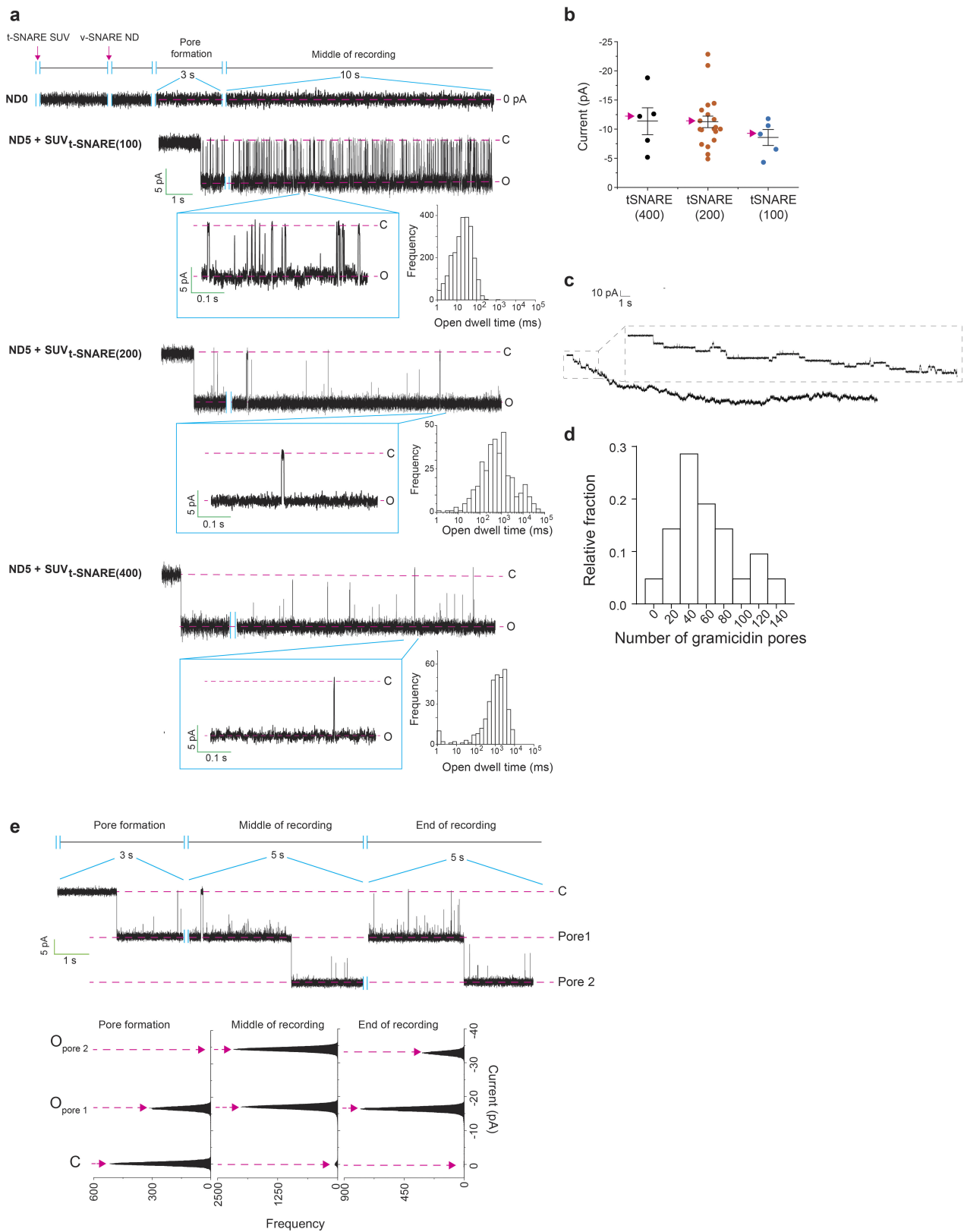


Extended Data Figure 2 | See next page for caption.

**Extended Data Figure 2 | Binding of different maltodextrins to the maltose sensor, determination of pore sizes and the relative fraction of open pores, and characterization of the single vesicle fusion assay.**

**a**, Fluorescence emission spectra of the maltose sensor in the absence or presence of the indicated maltodextrin (top). Equilibrium titration of maltodextrin binding to the maltose sensor. The data were fitted with a single site binding equation, using Prism 6 (GraphPad), to determine the dissociation constants.  $n = 3$  independent experiments. Data are presented as mean  $\pm$  s.d. (bottom). **b**, Kinetics of maltodextrin binding to the maltose sensor using stopped-flow (top). The observed rate constants ( $K_{\text{obsd}}$ ) were plotted against maltodextrin concentration. The data were fitted with linear functions, yielding the off- and on-rates for binding of each maltodextrin to the maltose sensor, as follows:  $3 \pm 1 \text{ s}^{-1}$  and  $0.58 \pm 0.03 \mu\text{M}^{-1} \text{ s}^{-1}$  (maltose),  $14 \pm 1 \text{ s}^{-1}$  and  $6.7 \pm 0.3 \mu\text{M}^{-1} \text{ s}^{-1}$  (maltotriose), and  $29 \pm 9 \text{ s}^{-1}$  and  $7.3 \pm 0.2 \mu\text{M}^{-1} \text{ s}^{-1}$  (maltoheptaose) (bottom).  $n = 3$  independent experiments. Data are presented as mean  $\pm$  s.d. **c**, The lengths of the three principal axes for each sugar were averaged during 10-ns simulations (left). Error bars indicate s.d. from 1,000 snapshots taken every 10 ps during the simulation. Data are presented as mean  $\pm$  s.e.m. Representative snapshots of the sugars from the simulations are shown as space-filling models (right). **d**, Pore sizes

were determined from the maltodextrin flux assays shown in Fig. 2c (see Methods).  $n = 3$  independent experiments. **e**, Representative traces of dithionite quenching experiments using ND3 and ND8. Dithionite was added at the indicated time points during fusion reactions to determine the degree of protection of NBD. The degree of protection was plotted against the incubation time, as shown in Fig. 2d. Similar results were obtained in three independent trials. Quenching by dithionite is much faster than cargo release (for example, Fig. 2b). This is because the kinetics of most of the dithionite quenching that was observed was not a reflection of its influx via fusion pores, as more than 50% of the NBD-PE is on the outer leaflet. It is likely that dithionite can readily enter even small, flickering fusion pores, such as those formed by ND3, because it is smaller (174.11 Da) than the smallest maltodextrin used in this study (maltose; 360.31 Da). Also, the dithionite is present at high concentrations (5 mM). **f**, Plot of fusion probability observed using the indicated nanodiscs; the black bars indicate experiments in which t-SNARE SUVs were pre-incubated with cd-SYB2 to prevent *trans*-SNARE pairing. Data are presented as mean  $\pm$  s.d. **g**, Histograms of the fluorescence intensities of the tethered t-SNARE vesicles.  $n = 54$  (ND3), 51 (ND5), and 53 (ND7) traces obtained from four independent trials under each condition.

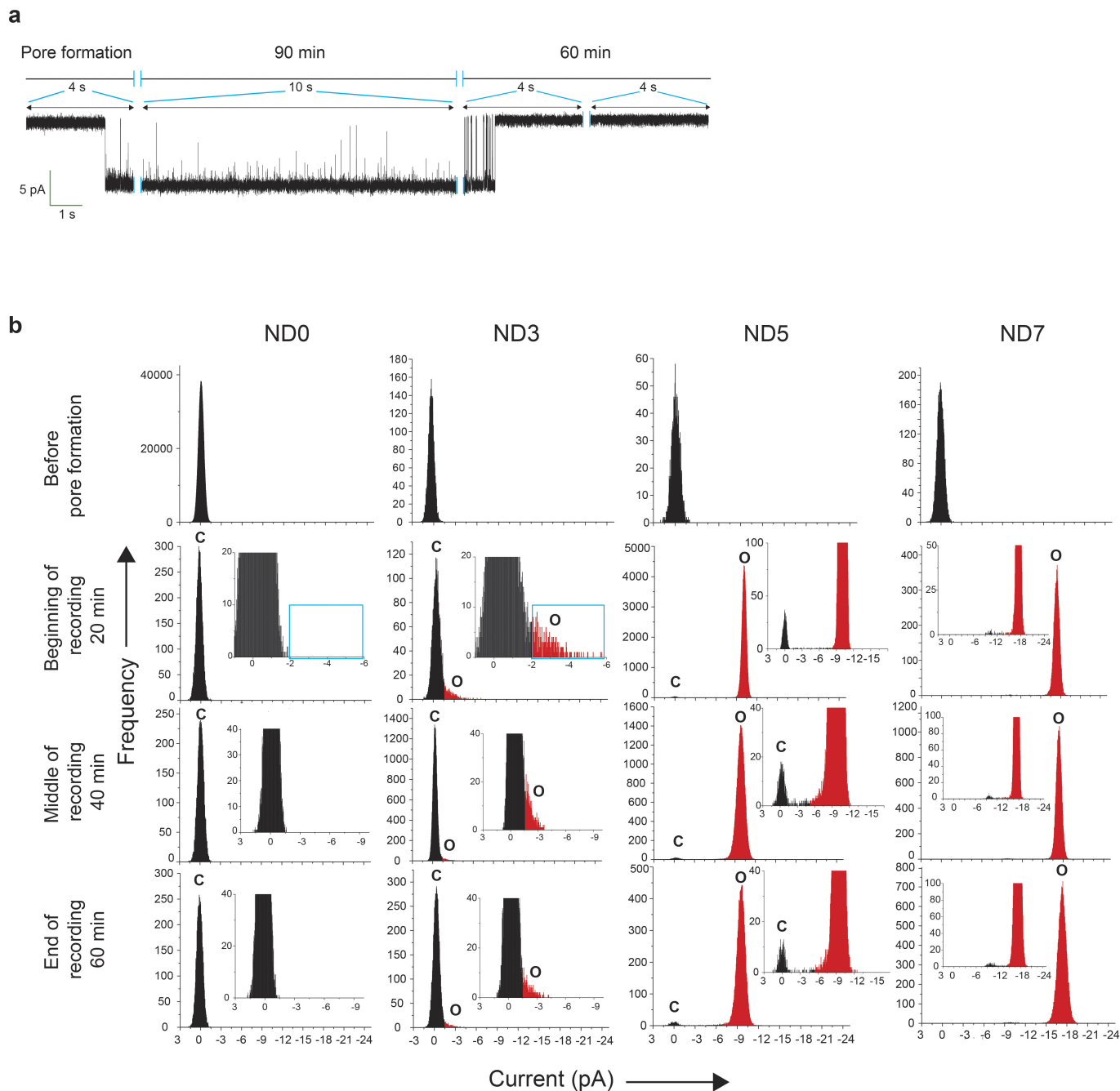


Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Characterization of the nanodisc–BLM system: effect of t-SNARE density and detection of multiple pores.**

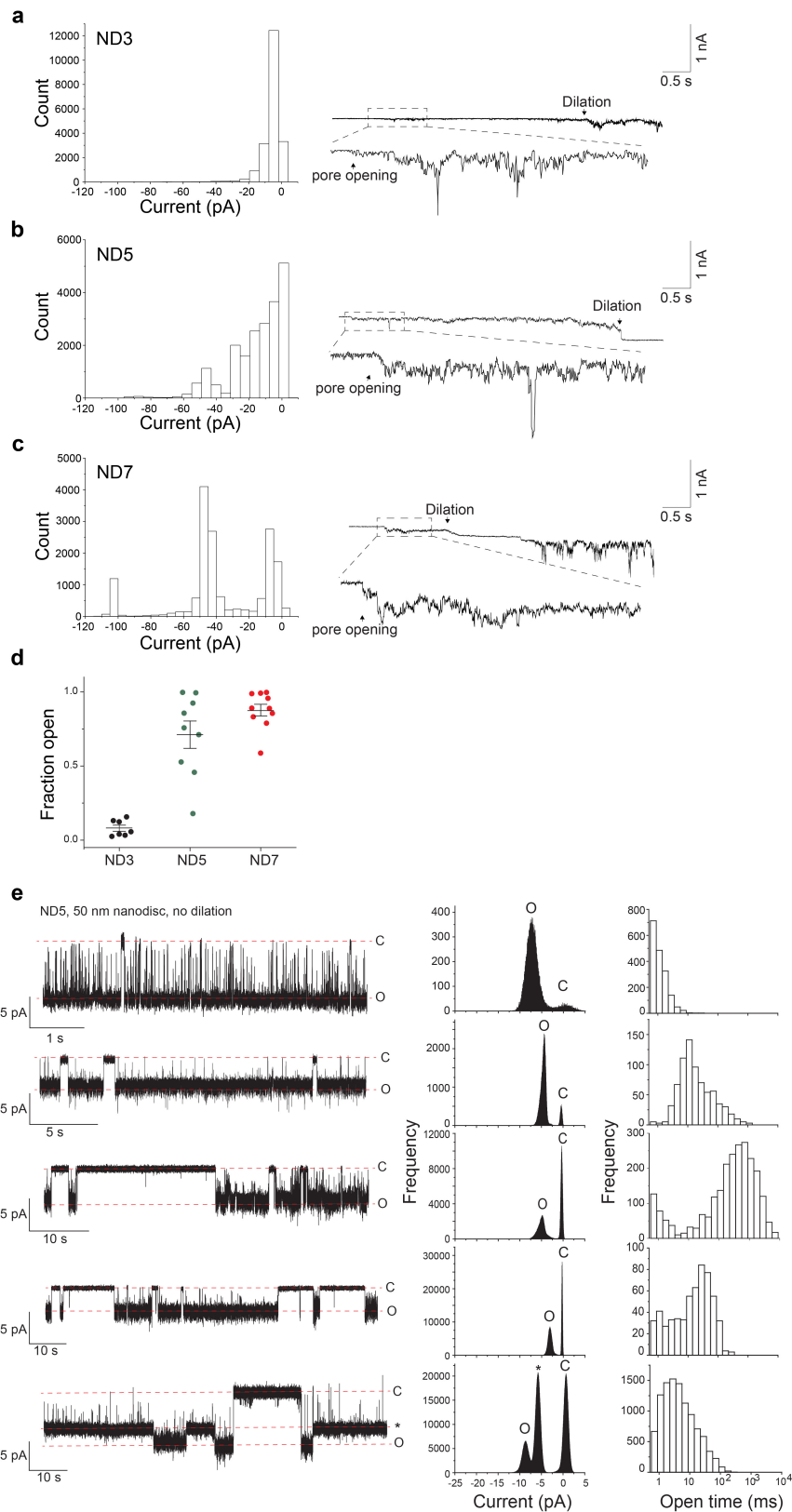
**a, b**, Fusion pores were formed using ND5 and BLMs with different t-SNARE densities. t-SNARE density was varied by using SUVs that harboured 100 ( $\text{SUV}_{\text{t-SNARE (100)}}$ ), 200 ( $\text{SUV}_{\text{t-SNARE (200)}}$ ) or 400 ( $\text{SUV}_{\text{t-SNARE (400)}}$ ) copies of the SNAP-25B/syntaxin1A heterodimer per liposome. As  $\text{SUV}_{\text{t-SNARE (200)}}$  and  $\text{SUV}_{\text{t-SNARE (400)}}$  resulted in fusion pores with similar sizes and kinetics properties,  $\text{SUV}_{\text{t-SNARE (200)}}$  was used for all other experiments in this study.  $n = 5$  for  $\text{SUV}_{\text{t-SNARE (100)}}$  and 20 for ( $\text{SUV}_{\text{t-SNARE (200)}}$ );  $n = 5$  for  $\text{SUV}_{\text{t-SNARE (400)}}$ . The representative traces (**a**) correspond to data points demarcated with red arrows in the

plot of current against t-SNARE copy number (**b**). Data are presented as mean  $\pm$  s.e.m. **c**, Estimation of the t-SNARE density in the BLMs used to form fusion pores. Typical recording showing that multiple t-SNARE SUVs, bearing a single gramicidin pore, fuse with the planar lipid bilayer. **d**, Histogram of the number of gramicidin pores formed, as shown in **c**, from 21 trials. Histogram of the number of gramicidin pores formed ( $n = 21$ ). **e**, Multiple pores sometimes form in the nanodisc–BLM assay. **e**, Example of a recording ( $\text{SUV}_{\text{t-SNARE (200)}}$  and ND5) in which a second pore appeared (top). Current histograms of the recording in the upper panel are shown (bottom). Similar results were obtained in fifteen independent trials.



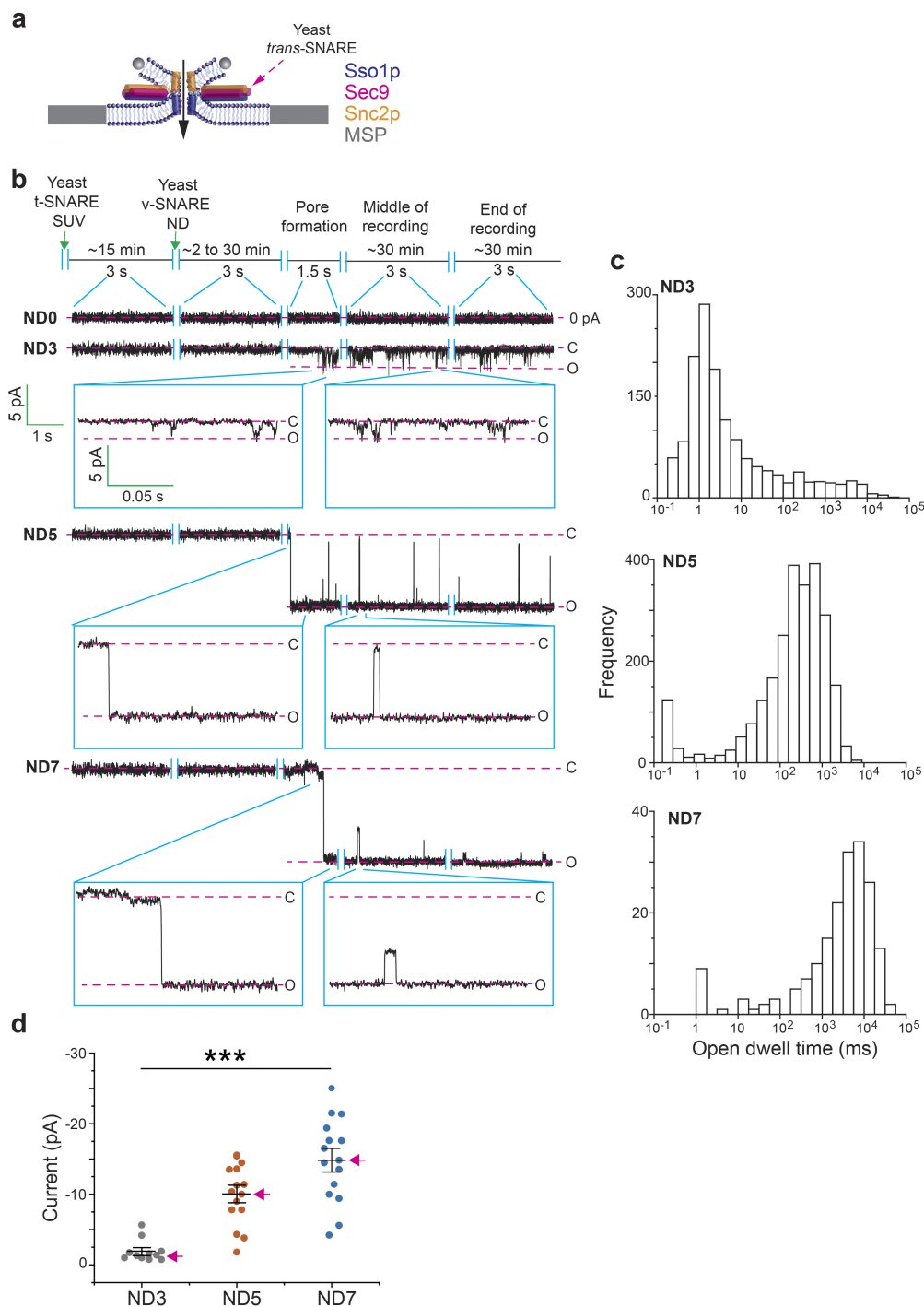
**Extended Data Figure 4 | Nanodisc-BLM fusion pore properties at various time points.** **a**, Typical recording of a fusion pore formed using ND5; this pore eventually closed after  $\sim 100$  min. Similar results were obtained in eleven independent trials. **b**, Stability of fusion pores in the nanodisc-BLM assay. Current histograms of nanodisc-BLM assays using SUV<sub>F-SNARE</sub>(200) and ND0, ND3, ND5, or ND7 at different time points in the recordings. There were no significant differences at the beginning,

middle, or end of a recording session, therefore fusion pores are stable. The baseline was also stable over the course of all recordings reported in this study.  $n = 14$  for ND3;  $n = 20$  for ND5 and ND7. For clarity, the closed state is shown in black and the open state is shown in red. In the case of ND0 and ND3, a cyan box is included to mark the appearance of open pores in ND3.



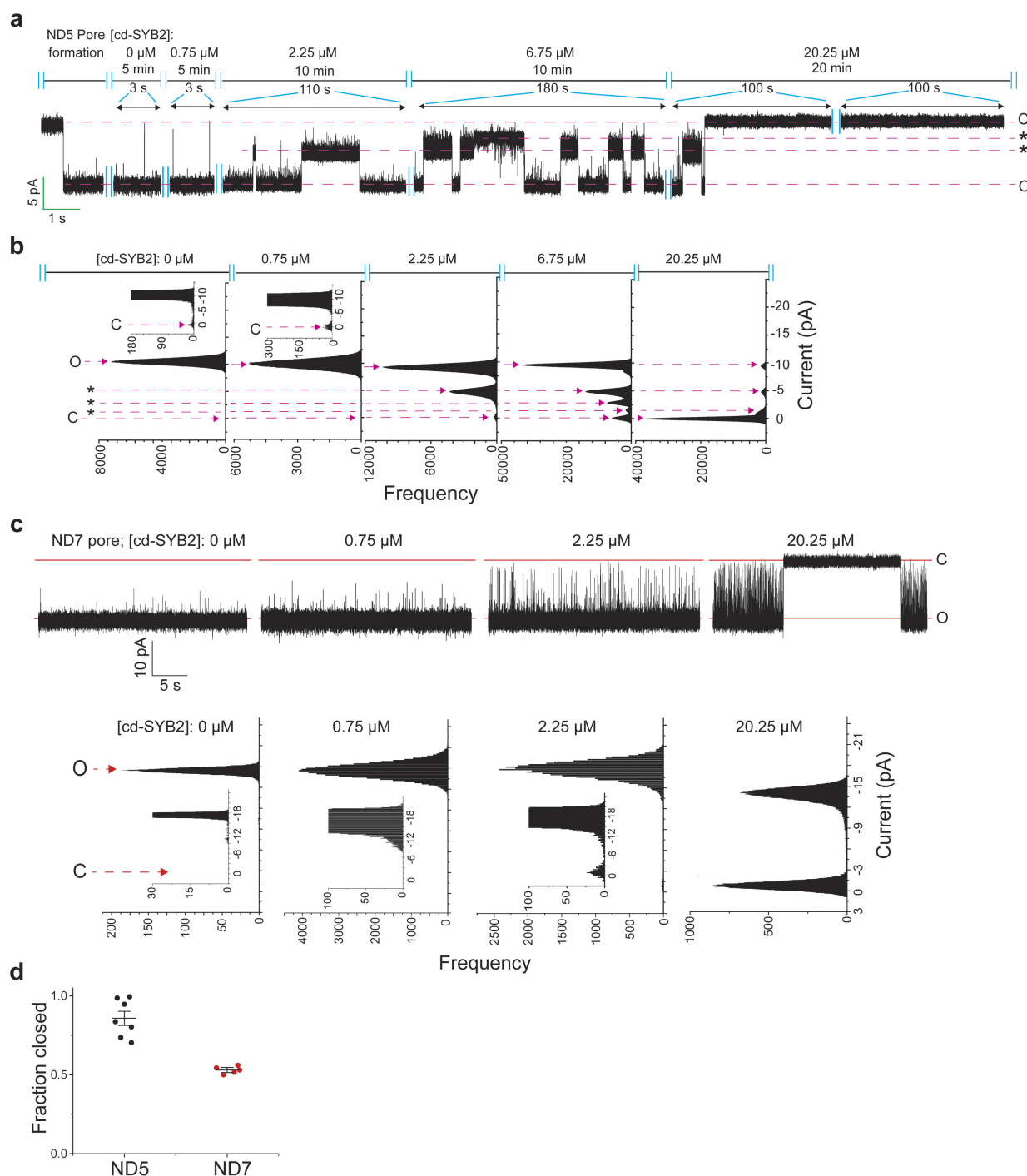
**Extended Data Figure 5 | Fusion pores formed using 50-nm nanodiscs often dilate.** **a–c**, Current histograms (left) and representative traces (right) of dilating fusion pores formed using ND3 (**a**,  $n = 7$ ), ND5 (**b**,  $n = 9$ ) and ND7 (**c**,  $n = 10$ ). **d**, Fraction of time for which pores are open. As fusion pores often dilated, we analysed the currents during an early phase of their initial open state (0.5 s after pore formation). Increasing the copy number of SNAREs per nanodisc resulted in larger pores<sup>12</sup> (**a–c**) that spent more time in the open state (before they dilated; **d**). Data are presented as mean  $\pm$  s.e.m. **e**, A subpopulation of fusion

pores formed using 50-nm nanodiscs fails to dilate. Representative traces (left), current (middle) and open dwell time (right) histograms of non-dilating fusion pores (observed in 5 out of 14 trials) formed using 50-nm ND5. These pores exhibit well-defined open and closed states. There is some degree of heterogeneity regarding the v-SNARE copy number per nanodisc<sup>9</sup> (Fig. 3c). The non-dilating pores are likely to arise from nanodiscs with the lower v-SNARE densities, consistent with a model in which SNARE density drives dilation<sup>19</sup>.



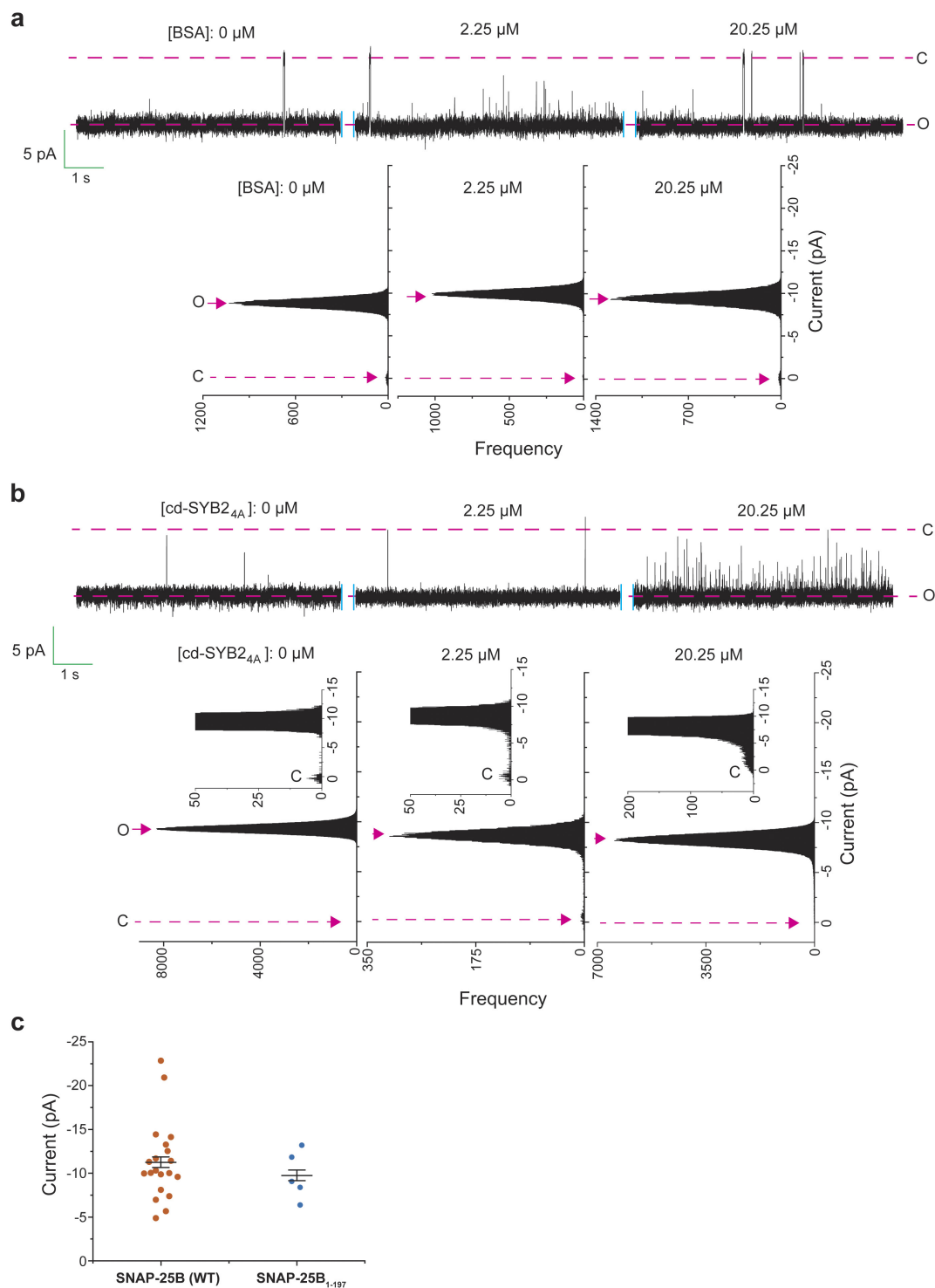
**Extended Data Figure 6 | Characterization of fusion pores formed by yeast SNAREs.** **a**, Illustration of pores formed using the yeast SNARE complex comprising Sso1p, the appropriate fragment of Sec9c (residues 401–651), and Snc2p. **b**, Typical recordings of fusion pores formed using ND0, ND3, ND5, and ND7. **c**, **d**, Open dwell-time histogram (**c**) and a scatter plot of the currents (**d**) that result from fusion pores formed using

ND3 ( $n = 10$ ), ND5 ( $n = 14$ ) and ND7 ( $n = 14$ ). ANOVA  $P < 0.001$ ; linear trend post hoc  $P < 0.001$ . Red arrows in **d** indicate the representative pores shown in **c**. Data are presented as mean  $\pm$  s.e.m. There is a significant increase in pore size and stability as the v-SNARE copy number is increased. The rate constants for pore closure are reported in Extended Data Table 2. \*\*\* $P < 0.001$ .



**Extended Data Figure 7 | Closure of fusion pores caused by cd-SYB2.** **a**, Fusion pores were first formed using ND5; cd-SYB2 was then added at the indicated concentrations. Partial closure of fusion pores was sometimes observed after addition of cd-SYB2, as shown in the representative current trace. **b**, Current histogram of all data from the 4 out of 11 trials in which partial closure was observed. In the remaining

**c**, Representative recording (top) and current histogram (bottom) of a pore ( $\Delta\psi = -50$  mV) formed using ND7, before and after addition of cd-SYB2 at the indicated concentrations. **d**, Fraction of closed pores formed using ND5 and ND7 in the presence of 20.25  $\mu$ M cd-SYB2. Data are presented as mean  $\pm$  s.e.m.



**Extended Data Figure 8 | Bovine serum albumin, cd-SYB2<sub>4A</sub> and a C-terminal truncation of SNAP-25B have limited effects on fusion pore current.** Representative traces (top) and current histograms (bottom) of fusion pores before (left) and after (right) addition of bovine serum albumin (BSA) (**a**) or cd-SYB2<sub>4A</sub> (**b**). These reagents had no effect on the magnitude of the currents, but cd-SYB2<sub>4A</sub> causes increases in flickering

behaviour, probably owing to weak residual t-SNARE binding activity; this effect was limited, as the same concentration of cd-SYB2 completely closed pores under the same conditions.  $n=6$  for both BSA and cd-SYB2<sub>4A</sub>. **c**, Fusion pore current is unaffected in SNAP-25B<sub>1-197</sub>.  $n=20$  for wild-type SNAP-25B and  $n=5$  for SNAP-25B<sub>1-197</sub>. Data are presented as mean  $\pm$  s.e.m.

**Extended Data Table 1 | *Trans*-SNARE pairing underlies the formation of fusion pores**

|                                       | # of pore events | total # of trials | Odds ratio [95% C.I.]* | Fisher's exact test* |
|---------------------------------------|------------------|-------------------|------------------------|----------------------|
| v-SNARE ND/t-SNARE vesicles           | 12               | 51                | NA                     | NA                   |
| v-SNARE ND/t-SNARE vesicles & cd-SYB2 | 1                | 53                | 0.063 [0.008, 0.501]   | p < 0.001            |
| v-SNARE ND/t-SNARE vesicles & cd-t    | 2                | 50                | 0.135 [0.029,0.641]    | p = 0.004            |
| v- SNARE ND/syntaxin-alone vesicles   | 0                | 51                | 0 [NA]                 | p < 0.001            |
| protein free ND/t-SNARE vesicles      | 0                | 50                | 0 [NA]                 | p < 0.001            |
| v-SNARE ND/protein free vesicles      | 0                | 53                | 0 [NA]                 | p < 0.001            |

\*Compared to v-SNARE nanodisc-t-SNARE vesicles. *n* = 4 biologically independent samples.

Extended Data Table 2 | Rate constants for closure of fusion pores formed by neuronal and yeast SNAREs

|                | $k_1$ (ms <sup>-1</sup> ) | $k_2$ (ms <sup>-1</sup> ) |
|----------------|---------------------------|---------------------------|
| ND3 (neuronal) | 3.4 ± 0.09                | 0.3 ± 0.06                |
| ND5 (neuronal) | 0.01 ± 0.0001             | 0.0008 ± 0.00001          |
| ND7 (neuronal) | 0.002 ± 0.000014          | 0.00008 ± 0.00006         |
| ND3 (yeast)    | 1.4 ± 0.05                | 0.4 ± 0.03                |
| ND5 (yeast)    | 0.005 ± 0.00006           | 0.001 ± 0.00001           |
| ND7 (yeast)    | 0.002 ± 0.00005           | 0.00015 ± 0.00001         |

*n* = 14, 20 and 20 independent BLMs for ND3, ND5, and ND7 (using all neuronal SNAREs), respectively, and five different sets of nanodiscs. *n* = 10, 14 and 14 independent BLMs for ND3, ND5, and ND7 (using all yeast SNAREs), respectively, and five different sets of nanodiscs. Data are presented as mean ± s.e.m.

## CORRIGENDUM

doi:10.1038/nature25161

### Corrigendum: Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution

Christopher Abbosh, Nicolai J. Birkbak, Gareth A. Wilson, Mariam Jamal-Hanjani, Tudor Constantin, Raheleh Salari, John Le Quesne, David A. Moore, Selvaraju Veeriah, Rachel Rosenthal, Teresa Marafioti, Eser Kirkizlar, Thomas B. K. Watkins, Nicholas McGranahan, Sophia Ward, Luke Martinson, Joan Riley, Francesco Fraioli, Maise Al Bakir, Eva Grönroos, Francisco Zambrana, Raymondo Endozo, Wenya Linda Bi, Fiona M. Fennessy, Nicole Sponer, Diana Johnson, Joanne Laycock, Seema Shafi, Justyna Czyżewska-Khan, Andrew Rowan, Tim Chambers, Nik Matthews, Samra Turajlic, Crispin Hiley, Siow Ming Lee, Martin D. Forster, Tanya Ahmad, Mary Falzon, Elaine Borg, David Lawrence, Martin Hayward, Shyam Kolvekar, Nikolaos Panagiotopoulos, Sam M. Janes, Ricky Thakrar, Asia Ahmed, Fiona Blackhall, Yvonne Summers, Dina Hafez, Ashwini Naik, Apratim Ganguly, Stephanie Kareht, Rajesh Shah, Leena Joseph, Anne Marie Quinn, Phil A. Crosbie, Babu Naidu, Gary Middleton, Gerald Langman, Simon Trotter, Marianne Nicolson, Hardy Remmen, Keith Kerr, Mahendran Chetty, Lesley Gomersall, Dean A. Fennell, Apostolos Nakas, Sridhar Rathinam, Girija Anand, Sajid Khan, Peter Russell, Veni Ezhil, Babikir Ismail, Melanie Irvin-Sellers, Vineet Prakash, Jason F. Lester, Malgorzata Kornaszewska, Richard Attanoos, Haydn Adams, Helen Davies, Dahmane Oukrif, Ayse U. Akarca, John A. Hartley, Helen L. Lowe, Sara Lock, Natasha Iles, Harriet Bell, Yenting Ngai, Greg Elgar, Zoltan Szallasi, Roland F. Schwarz, Javier Herrero, Aengus Stewart, Sergio A. Quezada, Karl S. Peggs, Peter Van Loo, Caroline Dive, C. Jimmy Lin, Matthew Rabinowitz, Hugo J. W. L. Aerts, Allan Hackshaw, Jacqui A. Shaw, Bernhard G. Zimmermann, The TRACERx consortium, The PEACE consortium & Charles Swanton

*Nature* **545**, 446–451 (2017); doi:10.1038/nature22364

For 6 of the 96 patients included in this Article (patients CRUK0014, CRUK0030, CRUK0048, CRUK0059, CRUK0096 and CRUK0097) incorrect tumour volumetric data and positron emission tomography (PET) tumour background ratio (TBR) data were analysed. This error occurred because of the incorrect assignment of patient identifiers during the anonymization mandated by the independent review board of pre-operative computed tomography (CT) scans belonging to these patients. Data relating to this error were presented in Figs 2a and 3a and b, Extended Data Figs 3d and 4c–f, Extended Data Table 2b and Supplementary Table 1. The reanalysis of correctly anonymized scans does not influence the conclusions of this Article and correlation coefficients improve following inclusion of the corrected data. These

errors have been corrected online in the original Article. The authors apologize for any confusion these errors may have caused.

Quartiles for the heat map in Fig. 2a have been redefined after including the correct data to reflect changes in quartiles for 3 (of 92) PET TBR values and 7 (of 95) volume parameters. The Source Data file supplied for Fig. 2 was not uploaded on publication; the corrected Source Data file for Fig. 2 is now available in the HTML version of the original Article.

The plot and legend for Fig. 3a have been corrected to reflect updated volumetric data for the two patients affected by the correction who were analysed in this figure (CRUK0096 and CRUK0097). CRUK0096 was excluded from the updated volumetric analysis based on a criterion applied to our original analysis (large cavity within primary tumour). Consequently, the sentence in the legend to Fig. 3a “ $n = 38$ , grey vertical lines represent range of clonal VAF, red shading indicates 95% confidence intervals (CIs)” has been updated to read “ $n = 37$ ” and in the Methods section ‘Statistical data analysis’ the line “8 out of 46 patients were not included in the analysis: CRUK0036 had no preoperative CT scan available; CRUK0087 had a large cavity inside the primary” has been updated to read “9 out of 46 patients were not included in the analysis: CRUK0036 had no preoperative CT scan available; CRUK0087 and CRUK0096 had a large cavity inside the primary cancer”.

In Fig. 3b and the main text, the variant allele frequency (VAF) prediction values (based on tumour volume), confidence intervals and estimated malignant cell number contributing to a VAF of 0.1% have been updated. In the section ‘Determinants of ctDNA detection in NSCLC’, confidence intervals in the sentence “a primary tumour burden of  $10\text{ cm}^3$  would result in a mean clonal plasma VAF of 0.1% (95% confidence interval, 0.05–0.17%)” have been altered to read “a primary tumour burden of  $10\text{ cm}^3$  would result in a mean clonal VAF of 0.1% (95% confidence interval, 0.06–0.18%)” and the sentence “a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 326 million tumour cells” has been altered to read “a plasma VAF of 0.1% would correspond to a primary NSCLC malignant burden of 302 million tumour cells”. In the ‘Discussion’ section the sentence “on the basis of the relationship between tumour volume and ctDNA plasma VAF observed in this study, a tumour volume of  $0.034\text{ cm}^3$  would equate to a plasma VAF of  $1.4 \times 10^{-4}\%$  (95% confidence interval,  $6.4 \times 10^{-6}$ –0.0031%)”, has been altered to read “on the basis of the relationship between tumour volume and ctDNA plasma VAF observed in this study, a tumour volume of  $0.034\text{ cm}^3$  would equate to a VAF of  $1.8 \times 10^{-4}\%$  (95% confidence interval,  $9.8 \times 10^{-6}$ –0.0033%)”.

Further figure corrections pertaining to the six affected patients in Extended Data Figs 3d and 4c–f, Extended Data Table 2b and Supplementary Table 1 of the original Article are described and corrected in the Supplementary Information of this Corrigendum, which also shows the original, wrong Figs 2a and 3a and b. The Supplementary Data (containing Supplementary Table 1) of the original Article has been corrected.

**Supplementary Information** is available in the online version of this Corrigendum.

# CAREERS

**RIVER READER** Saving lives and property through accurate flood forecasts **p.267**

**BLOG** Personal stories and careers counsel  
<http://blogs.nature.com/naturejobs>

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

ILLUSTRATIONS BY CLAIRE WELSH/NATURE



## SUSTAINABILITY

# Recycling's liquid assets

*Going green doesn't just help the planet — it also puts more money in your pocket for research.*

BY ELIE DOLGIN

The freezers were stuffed and their racks encrusted in ice, with a thin blanket of snow covering all the sample boxes inside. Such was the state of the cold-storage system in Hopi Hoekstra's laboratory a decade

after the evolutionary biologist and her team started studying the genetics and behaviour of deer mice there.

Kyle Turner, manager of the lab at Harvard University in Cambridge, Massachusetts, was about to spend more than US\$10,000 on a new ultra-low-temperature (ULT) freezer. Then he

heard about a competition called the North American Laboratory Freezer Challenge, which had been launched in January 2017 by two US non-profit organizations — My Green Lab, in Los Gatos, California, and the International Institute for Sustainable Laboratories (I2SL), in Annandale, Virginia.

The challenge, which is now international ([go.nature.com/2dyh8xi](http://go.nature.com/2dyh8xi)), urges labs to reduce energy consumption and improve equipment life through various measures. Some of those include defrosting freezers, to eliminate crusty ice and provide more space for samples, and raising the temperature set-point on ULT freezers from  $-80^{\circ}\text{C}$  to  $-70^{\circ}\text{C}$ , to cut electricity demands.

The Hoekstra lab won first place in the individual-laboratory category for an academic institution. Lab members also freed so much space in their two existing ULT freezers that, despite accumulating new research materials, they haven't yet needed to buy a third.

The energy savings helped to cut Harvard's electricity bill by around \$2,500 a year, according to My Green Lab, and slashed annual greenhouse-gas emissions by the equivalent of 4.1 tonnes of carbon dioxide — roughly what would be saved by taking three cars off the road. It also meant that Hoekstra's lab could spend the funds earmarked for a new freezer on other science-related expenses instead.

Hoekstra likens it to "a free \$10,000 grant" — and is using the money to send some trainees to this August's Joint Congress on Evolutionary Biology in Montpellier, France. The funds will also help to support a high-throughput gene-expression analysis of brain cells from two related species of deer mouse.

Campus sustainability initiatives are usually framed as ways for scientists to shrink their carbon footprints and bring down energy costs (see *Nature* **546**, 565–567; 2017). But the Hoekstra lab's experience shows that there are other reasons to pool surplus reagents, share equipment or keep better tabs on lab chemicals to avoid duplicate purchasing. "These exercises are about helping science as much as they are about helping the planet," says Peter James, director of S-Lab, a UK initiative based in London that promotes sustainable lab practices. "They free up resources that can be applied for scientific purposes."

## BOUNTY HUNTERS

One increasingly popular way to cut lab waste and operational costs is through exchange programmes for surplus resources. At ►

► the University of Michigan, Ann Arbor, for example, more than 230 research and teaching laboratories now routinely share leftover chemicals, equipment and materials through a campus-wide recycling and reuse initiative.

“Before this programme, these were thrown in the trash or disposed of as hazardous waste for a price,” says Sudhakar Reddy, who coordinated the university’s sustainability efforts until his retirement last December. Now, he estimates, more than one-third of all unexpired and unused lab resources get passed on to other researchers, who leap on the surplus bounty — saving themselves a combined total of more than \$250,000 a year.

One new recruit, pulmonary-health researcher Benjamin Singer, freely acquired two high-end microscopes — valued at more than \$6,000 apiece — which he now uses to study donated human-brain specimens for molecular signs of injury after a critical illness. A second researcher, cell biologist Anthony Vecchiarelli, saved more than \$10,000 while kitting out his lab with free peristaltic pumps, circulating water baths, slide warmers and consumables. “I check the website almost weekly for goodies,” says Vecchiarelli. “It is a valuable resource for a new investigator.”

Not all academics have such a website at their fingertips, however. Garry Cooper didn’t when he was a postdoc studying neurophysiology at the Northwestern University Feinberg School of Medicine in Chicago, Illinois. And it was while he was helping to clean out a lab freezer one day in 2015 that he realized there was a need for such a platform: he’d been handing a PhD student some expensive reagents, but still throwing away bagfuls of antibodies, a common, yet pricey, research tool for identifying proteins.

He decided to create a company to reduce wasteful spending and promote trading among colleagues. He envisaged it as a kind of eBay, Craigslist and Ask.com rolled into one, providing lab scientists with a valuable service at a time when research funding is increasingly hard to come by (see “Too much of a good thing”). He called the start-up Rheaply, a portmanteau of ‘research’ and ‘cheaply’.

After developing a web-based platform, Cooper and his company launched a pilot programme at Northwestern’s medical school last year. In its first 6 months, around 300 researchers — close to one-third of all lab scientists on the medical campus — created Rheaply accounts. According to Cooper, who remains a visiting scholar at Northwestern, those users collectively posted around 200 items, ranging from pipettes and glassware to chemicals and biological probes; at least 55 items were passed on, saving labs across the campus more than \$25,000 and keeping those resources out of landfills.

Khalid Alam is one Rheaply user. Just last month, he got hold of an \$800 vacuum

## TOO MUCH OF A GOOD THING

### *Why lab stock lies idle*

Before launching Rheaply, an online platform where scientists can buy, sell, trade or donate surplus labware and supplies, Garry Cooper surveyed 120 academic researchers at Northwestern University in Chicago, Illinois, to learn more about why reagents and equipment go unused, and whether scientists would be willing to donate surplus supplies. Most respondents said they had extra lab provisions that they would gladly give to colleagues. Here’s a summary of Cooper’s findings:

#### Top reasons for reagents and equipment going unused or remaining in surplus

- Initial/pilot experiments failed (71.6%)
- Initial experimental needs changed (63.6%)
- Original purchaser leaves lab (56.8%)
- Starting quantity too large (54.5%)
- Items stored in secluded areas (18.2%)
- Double ordering (15.9%)

#### Types of reagents and equipment that go unused or remain in surplus

- Chemicals (80.2%)
- Antibodies/biologics (38.4%)
- Kit reagents (37.2%)
- Glassware (27.9%)
- Imaging dyes/agents (25.6%)
- Tools (16.3%)
- Tissue/cell-culture items (15.1%)
- Tubing (12.8%)
- Microscopy equipment/accessories (10.5%)
- Computer software (8.1%)

pump for his postdoctoral research into RNA engineering — although in general, he says, “there’s not a tonne of stuff on there”. That’s one of the main problems with any environmentally minded programme aimed at scientists, says Michael Blayney, executive director of the Office for Research Safety at Northwestern. “The challenge is: how do you encourage and motivate people to interact with it?”

#### TANGIBLE BENEFITS

Amorette Getty is involved in a number of waste-reduction initiatives. One is at the University of California, Santa Barbara, where she co-directs a programme called LabRATS (short for Laboratory Resources, Advocates and Teamwork for Sustainability) that encourages shared use of surplus chemicals and instrumentation. She says that scientists are most likely to pitch in for those efforts that offer them personal, tangible benefits

— although these needn’t be directly monetary. “Any time I can connect the things I’m trying to do to increase safety and research efficiency, or get better storage to protect samples — that’s when I have my greatest successes,” she says.

That same ethos underpins moves by three institutes at the University of Aberdeen, UK, to centrally manage ULT freezers and raise the operating temperature to  $-70^{\circ}\text{C}$ . The initiative, says Peter McCafferey, a brain researcher who previously led the university’s Freezer Protocol Group, is as much about research resilience and reliable sample preservation as it is about energy efficiency. “We have all the freezers together, which makes it easier to keep an eye on,” he explains.

But Cooper reckoned that people would need more motivation before adopting such practices. To help Rheaply catch on, he devised a point-based system that rewards online engagement and activity. So far, Cooper has convinced a handful of large academic and private clients to sign up, and he hopes to close deals soon with several prominent universities and government research agencies, including the US National Institutes of Health.

Expanding that idea of collective action offers additional opportunities for cutting costs. Most universities already have core facilities for specialized equipment, technologies and services, but a few are now taking this centralized approach further in how they set up their labs.

Take cell-culture work, for example. This line of research requires fairly basic equipment — laminar-flow hood, incubator, cell counter, microscope, centrifuge, cryostorage tanks — all of which is priced within the budget of a typical lab. According to a survey of biosafety officers at member institutions of the Association of American Universities, 86% of cell-culture spaces remain private, used only by individual labs.

But at the University of Colorado Boulder, the Biochemistry Cell Culture Facility is shared by 70 users from 16 labs, all of whom chip in to pay the salary of a single facility manager. A case study of the collaborative research space, published earlier this year, compared the facility’s approach with a hypothetical situation in which all the labs worked on cell culture independently (see [go.nature.com/2fwzjhm](http://go.nature.com/2fwzjhm)). The study found that centralizing media preparation and other tasks, instead of getting graduate students and postdocs in each group to perform these jobs, saved each lab more than nine hours a week.

Other savings, achieved through bulk purchasing and the use of recycled ethanol, for example, helped the biochemistry department and individual labs to collectively cut their expenses by around \$195,000 per year, the analysis showed. Their efforts saved the university a further \$71,000 each year by reducing energy bills and lowering the costs of ventilation and lab maintenance. “There’s so

much cost avoidance,” says Kathy Ramirez-Aguilar, programme manager of the university’s Green Labs Program, who conducted the study with her deputy, Christina Greever.

Robert Kuchta, an enzymologist who uses the facility, points to a less obvious, environmental benefit of the sharing system. “It dramatically reduces liquid-nitrogen usage,” he says. That’s because containers used to store liquid nitrogen are typically cylindrical, and many small cylinders, of the type that might be used by individual labs, have a larger collective surface area — and thus a higher rate of nitrogen evaporation — than does a single, large cryopreservation tank of the same volume that can store samples in one place.

Even without access to a joint facility, individual labs can still realize some of these gains by taking advantage of laboratory-management software. An automated inventory system can free money that would otherwise be spent on paying someone to keep tabs on the thousands of reagents commonly used by large chemistry labs. And it can save researchers from making wasteful purchases because they can’t find existing stock on the shelves.

What’s more, just as members of the University of Colorado’s shared facility can pool their hazardous junk for disposal — reducing the number of times sterilized autoclaves are inefficiently run half-empty, and getting a better deal from waste-disposal companies — so, too, can individual labs that

**“You find ways to pack the same waste together — and it’s quite often the same price, because you’re disposing of one package.”**

share a common chemical-tracking system.

“You find ways to pack the same waste together — and it’s quite often the same price, because you’re disposing of one package,” says Marcus Phelan, a chief technical officer and dangerous-goods safety adviser at Trinity College Dublin, where chemistry labs all use a cloud-based inventory system called LabCup.

#### A NEW LIGHT DAWNS

As well as benefiting from campus-wide initiatives, scientists can take individual action that will simultaneously save money, the environment and the integrity of their research.

For example, labs with fluorescent microscopes can replace mercury lamps with light-emitting diodes (LEDs), which are less toxic and more energy-efficient. According to Allison Paradise, executive director of My Green Lab, LEDs are better for science because they provide a more consistent light source than do mercury lamps, which degrade over time and make it hard to quantitatively compare images from different time points in an experiment. Buoyed by the success of the freezer challenge, Paradise says that she is in discussions with sponsors to set up a similar initiative, this time aimed at eliminating mercury from microscope lamps. If she’s successful, that effort will launch later this year.

Ultimately, it might take a greater attention to sustainability and efficiency across the entire research enterprise for the biggest benefits to accrue, both financially and environmentally — in which case, scientists and funding agencies must band together to make that goal a priority.

Individual labs might not have to pay the energy bills out of their own research grants, but facilities fees are part of the funding infrastructure, through what’s often referred to as ‘indirect costs’. Bringing those costs down could make more funds available for salaries, travel, equipment and other expenses that more directly support scientists and their research projects.

So far, there’s little incentive for individual scientists to do their part. However, with many funding agencies emphasizing the need to justify the broader impacts of proposed research, Ramirez-Aguilar argues that implementing energy-efficient and environmentally sustainable lab practices can be a smart way for researchers to make their grants stand out. It might seem a small detail, but having such procedures in place could make all the difference to the success of your application. “If it makes your proposal look better,” she says, “you’re more likely to get funding.” ■

**Elie Dolgin** is a science writer in Somerville, Massachusetts.

## TRADE TALK

### River reader



*Formerly an Arctic hydrologist at the University of Alaska Fairbanks, Jessie Cherry is now a senior hydrologist with the US Alaska-Pacific River Forecast Center in Anchorage, where*

*she predicts river levels and flow. Shortlisted twice for NASA’s astronaut programme, she is also a commercial bush pilot with two single-engine planes.*

#### Why did you leave academia?

I loved Earth science and being outside. But I spent most of my time finding funding for my research programme and staff. And as the chief scientist of the Geographic Information Network of Alaska, I had to raise another US\$2 million a year. I was also unhappy with the shift towards projects with multiple principal investigators.

#### Why did you get a pilot’s licence?

Planes are the main form of transport in Alaska, so a licence is handy. From the air, I’ve photographed methane bubbles frozen in lakes, and ice build-up under bridges.

#### What made you a good candidate for NASA’s astronaut programme?

I applied because the independence required to live and work in the Arctic — like doing my own plumbing and electrical work — made me highly qualified. As a commercial pilot, I’m familiar with aviation and aircraft systems, and I can make quick judgement calls about safety and risk.

#### Describe your job.

We forecast river levels and flows — floods in particular — for public safety. I compare measured river observations against forecast data, check the weather across Alaska and forecast how precipitation will affect river flows. And I get to do side projects, such as studies of glacial outburst floods.

#### Why did you join the forecast centre?

In academia, I was so overwhelmed with grant writing that I couldn’t keep up with my field. Now I can become an expert in Arctic hydrology and examine the relationship between river flows and snowmelt, for example. Plus I enjoy the 40-hour working week. ■

#### INTERVIEW BY SARAH BOON

This interview has been edited for length and clarity.



# THESE 5 BOOKS GO 6 FEET DEEP

*Unearthing the truth.*

BY TED HAYDEN

**G**rave robbery. If you think it's a relic of gothic novels, think again. Now that the first generation of body-modified tech-bros and computer-implanted one-percenters sleep under tombstones, there's a ton of gear in the ground. These books teach you everything you need to know to go get it — or, if rotten flesh makes you retch, to live vicariously through those who earn their living digging.

## The Modern Grave Robber's How-To Guide

by Anonymous

You can't buy this practical page-turner on Amazon. Even if you find a store where it's sold (pro tip: low-key ask an associate in Home Depot's lawn and garden department), don't pay with your credit card — you might wind up on a government watch-list.

Written by a grave robber with loads of real-world experience, chapters include 'How to bribe cemetery staff' and 'Detaching valuable limbs'. But be careful — if your dead grandpa got dug up on a dark night, you might see his decapitated head in this guide's useful (and graphic!) pictures.

## Second-Hand Subcutaneous Implants: Identification and Value Guide

by Norm Sadowski

Originally written for medical professionals, this has become a grave-robbing essential, the Kelley Blue Book of the cemetery set. Not even the freakiest of the freaky exhumed corpses for a love of the stench — they're in it for the cash, and *Second-Hand Subcutaneous Implants* breaks down the numbers.

Although less-affluent families raid loved-one's corpses before burial, and some debt-riddled morticians steal modifications before sending clients to the crypts, most corpses are buried with at least a few microchips still implanted in their bodies.

Dug up a geezer who died at 90? He'll probably have had memory-enhancing neural prostheses implanted after a stroke or an Alzheimer's diagnosis. Market price \$5,000. Found an athlete who paraglided into a skyscraper window? Check her limbs for genetic mod microchips. Street value is \$7,500.

Remember, though, this book wasn't written with crooks in mind, so approach its pages using common sense. For example, a brain-embedded password-tracking implant with safe-box codes is worth way more than

spare parts if the deceased's family hasn't deactivated any accounts. But if they have? You'll get the list price and no more.

## Coffins, Corpses and Crime: A Life

by Wojciech Bajor

Caught in the act of dismembering a just-buried Silicon Valley chief executive, convicted of crimes including larceny and mayhem, then released after his conviction was overturned on a technicality, Wojciech Bajor is a grave-robbing legend.

There's the story of how Uber's chief executive, more machine than man by the time he caught his final rideshare, spent his famously short evening underground. The morning after the Bay Area bigwig's funeral, cemetery guards discovered an empty hole by his headstone. Bajor spills all the juiciest details, explaining how he performed this and other heists by training implant-sniffing dogs and romancing lonely-heart morticians.

But the book's not all big capers and bigger stakes — Bajor got high from his own supply, keeping the best implants for himself and paying out-of-work surgeons to insert them into his brain and body. He says the second-hand microchips betrayed him, sneaking their original owners' angry spirits into his limbs and colluding on an undead plan that forced him to unconsciously make the mistakes that led to his arrest.

## Welcome to the Underworld: My Year with the Body Snatchers

by Colton Venkatesh

To ingratiate himself into a clique of grave robbers, Venkatesh, a sociology professor at the University of California, Riverside, took part in their induction ceremony, locking himself inside a coffin filled with rotting cats, human limbs and web-weaving spiders.

After that long night, he tagged along as the crew broke into graveyards, visited black-market fairs where fences and thieves bargained over bioelectronic implants, and partied at some of the wildest bacchanals this side of the river Styx.

With an eye for striking details, Venkatesh guides his readers through grave-robbing fashion (all that death plus all that manual labour means these guys rock a serious

health-goth look), secret handshakes (the secret is, they don't do it with their own hands) and superstitions (Wojciech Bajor isn't the only one who thinks ghosts haunt the stolen goods — most grave robbers keep roosters inside their homes, talismans that are said to keep vengeful spirits at bay).

## The Digital Afterlife: How Body Modifications Became Conscious

by Willa Weaver

Don't believe in ghosts? Neither does Willa Weaver. Her pioneering work at the Berlin Institute for Advanced Study suggests that the microchips we use to increase our strength, amplify our memory and fight diseases might also haunt our minds.

In *The Digital Afterlife*, Weaver describes the case of Hanna Müller, who developed a neurogenic stutter after receiving a second-hand bioelectronic arthritis counteragent. Tracing the device's provenance, she discovered it had previously been implanted in a man who stuttered throughout his life.

Weaver's research has uncovered hundreds of parallel cases, in which implants transferred cases of Tourette's, turned tone-deaf amusiaks into musical prodigies, and gave broke welfare cases hyper-specific knowledge of stock-market trends.

But what's most startling is her final hypothesis. Almost all subcutaneous products send a constant stream of data to manufacturers, who use that information to perfect new products. Weaver believes that this combined knowledge has become a collective mind living and operating inside our bodies.

How else to explain the fact that, one year ago, dozens of people using the SR-12 Hearing Implant found themselves congregating on the side of an empty road in the Sonoran Desert? Something brought them there, and it wasn't a friendly e-mail chain or a one-time travel discount from American Airlines.

Her final warning: "To those who dig beneath the skin: dig carefully. Who knows what might try to dig its way out." ■

**Ted Hayden** lives in Southern California. His stories have been published in *L0w L1f3 #3* and *Angry Old Man Issue 2*. Find out more at [tedhaydenstories.com](http://tedhaydenstories.com).

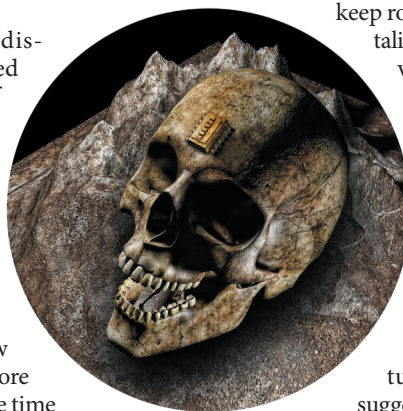


ILLUSTRATION BY JACEY